"Estimation of the efficacy of seasonal influenza vaccine : vaccine development, statistical modelling and trial design"

Benoît, Anne

Abstract

Influenza is an infectious seasonal disease against which yearly vaccination is recommended, especially for high risk populations. For a new vaccine, efficacy is classically assessed in large phase III trials. Unfortunately, in the past years, many trials have led to unexpected results. This work first aims to identify the particularities of the context of influenza with regards to the vaccine development. We confirm the link between post vaccination antibody response and vaccine efficacy without being able to identify an absolute threshold of protection. Combining simulations and analytical results, we then substantiate the limits of the actual statistical regression models used to estimate vaccine efficacy. To reach this goal, we had to develop a simulation framework for the generation of phase III clinical trials time-to-infection data. Our methodology brings together particularities of influenza and other infectious diseases as well as historical data. Finally, we propose a new m...

Document type : *Thèse (Dissertation)*

Estimation of the efficacy of seasonal influenza vaccines:

Vaccine development, statistical modelling and trial design

A thesis submitted in partial fulfillment of the requirements for the degree of Docteur en Sciences, Université catholique de Louvain, January 2015

ANNE BENOIT

Thesis committee:

Catherine Legrand, Supervisor (Université catholique de Louvain)
Pierre Devolder, President (Université catholique de Louvain)
Walthère Dewé (GSK Vaccines)
Philippe Lambert (University of Liège, Université catholique de Louvain)
Niko Speybroeck (Université catholique de Louvain)
Fabian Tibaldi (GSK Vaccines)

"Perfection is a theory. You cannot be a perfect human being, perfect artist. But go through your daily routine with hope you will be a little better in all respects, and do something meaningful" – Mikhail Baryshnikov

Acknowledgments

It is a pleasure to thank the many people who made this thesis possible.

I would like to express my special appreciation and thanks to my advisor, Catherine Legrand. Catherine, thank your for your support and guidance throughout this project. You have helped me to grow as a research scientist and were always there to encourage me when I needed it.

I would like to thank the members of my committee, Walthère Dewé, Philippe Lambert and Niko Speybroeck. Walthère, thank you for entrusting me with this project, for your help and all the guidance you provided along the way. Thank you Philippe and Niko. You brought diverse talents and perspectives to the table that improved my project. I thank the jury members, Fabian Tibaldi and Pierre Devolder for their interest and their suggestions during my private defense.

I am grateful to all the administrative staff of ISBA, Nancy Guillaume, Sophie Malali, Nadja Pfeiffer, Maguy Hanon and Tatiana Regout. Beside the very much appreciated practical support, you provide this institute with a sense of family that I am gonna miss.

I would like to thank the people at GSK, for their help, suggestions and for providing valuable data to help develop and illustrate this thesis.

Thank you Pierre Lebrun, Vincent Bremhost and Cédric Taverne for the precious Bayesian help.

I would like to thank my colleagues whom I loved spending the past four years. First, I thank Marco, the best colleague ever! I also thank Adrien, Alina, Diane, Federico, Habiba, Jennifer and Sylvie who have made sharing an office such a tremendous experience. Thank you Alain, Anna, Aurélie, Benjamin, Catherine, Cédric, Céline, Fabian, Maïlis, Mathieu, Michal, Nathalie, Nathan, Nicolas, Réjane, Vincent, who make ISBA a warm and fun place to work.

Self-confidence has never been my strong point and beside the people I've already thanked, I would like to express my gratitude to Bernadette Govaerts, Christian Ritter, Bruno Boulanger, Astrid Jullion and Tom Jacobs for believing in me and helping me believe in myself.

I have been blessed with a very loving and supportive family. Words cannot express how grateful I am to my parents, my brother and my sister for their love and encouragement. You have always believed in me! Special thanks to Jean-Luc, my American uncle, for the much appreciated English review. Thank you to my husband's family for their support and understanding.

The last person I want to thank is my husband. Olivier, thank you for being there, for encouraging and supporting me and most of all, for reminding me there is a whole world outside of my PhD.

Contents

List of abbreviations

AR ..... Attack rate

CDC ..... Centers for Disease Control and Prevention

CI ..... Confidence interval

CID ..... Communicable infectious disease

COP ..... Correlate of protection

CT ..... Clinical trial

EMEA ..... European Medicines Evaluation Agency

EPV ..... Events per variables

Exp. ..... Experimental vaccine

FDA ..... Food and Drug Administration

FN ..... False negative

FP ..... False positive

GMT ..... Geometric mean titre

HI ..... Antibodies against influenza Haemagglutinin

HR ..... Hazards ratio

IR ..... Incidences ratio

OECD ..... Organization for Economic Co-operation and Development

OR ..... Odds ratio

PCR ..... Polymerase Chain Reaction

PH . . . . . Proportional hazards

PI . . . . . Prediction interval

Ref. . . . . . Reference vaccine

ROC . . . . . Receiver Operating Characteristic

RR . . . . . Risk ratio

SOP . . . . . Surrogate of protection

TI . . . . . Tolerance interval

TN . . . . . True negative

TP . . . . . True positive

VE . . . . . Vaccine efficacy

WHO . . . . . World Health Organization

Chapter 1

Introduction

*

1.1 The context

Influenza virus infects from 5 to 10% of the population yearly (WHO 2012). Influenza causes a benign if uncomfortable illness is healthy adults, but can lead to sever complications, especially in fragile populations, including the elderly, young children and people with concomitant conditions. Indeed, influenza seasonal epidemics cause an estimated 250000 to 500000 deaths yearly (WHO 2012). Therefore, vaccination is recommended by the World Health Organisation (WHO) especially in the high risk populations. Unfortunately, ageing is associated with a decline in immune function which results in reduced vaccine efficacy (VE) (Haq and McElhaney 2014). For this reason, pharmaceutical companies are developing novel seasonal influenza vaccines that offer better protection to this population.

To obtain marketing licensure, development of a new human medicinal product goes through several clinical phases (ICH8 1998; WHO 2004). Phase I studies designed to evaluate the safety of a product. Phase II studies are designed to provide preliminary information about a a product to achieve a desired effect. In vaccine phase II studies, the key vaccine effect is the induction of potentially useful immune responses, such as the production of high levels of antibody against the targeted infection. Whether to grant the marketing license is based on large phase III trials, called pivotal studies. In a phase III trial, the efficacy and the safety of the product must be fully assessed. In vaccine phase III trials, the efficacy outcome of interest is the occurrence (or not) of the targeted disease, confirmed through laboratory analyses.

The clinical VE of a seasonal influenza vaccine is usually assessed in a large multicentered phase III trial. The objective of such a trial is to show that the experimental vaccine reduces the risk of influenza as compared to a reference (placebo or a standard vaccine) (Chan et al. 2003). Participants are vaccinated with either the experimental

or the reference vaccine prior to the influenza annual season. During the surveillance season, i.e. from November to April, influenza cases and times of onset are collected. In subjects who present influenza symptoms, such as fever and cough, a nasopharyngeal swab is collected and analysed to confirm influenza infection.

1.2 Motivation and objectives

In the last decades, many trials have led to unexpected results, such as failure to achieve any significant efficacy, or discordant results from one season to the next using the same vaccine candidate and study design (Beran et al. 2009a,b). Although lack of efficacy is unexpected, it could be the result of using an ineffective product. However, discordant results are more perplexing (Dewé et al. 2013).

Our project has two main objectives: the first objective is to understand the particularities of seasonal influenza and the context in which VE trials take place, by identifying the various factors and sources of heterogeneity which may lead to achieving opposite results with the same vaccine. The second objective is to use this information to improve the design of future efficacy trials, to propose a new way of thinking the design of future efficacy trials and to propose more appropriate statistical tools for the analyses of such trials. This dissertation is meant as a detective work. First it identifies the potential causes of failure of previous trials, then studies these hypotheses and tests their validity. Specifically, we will investigate four hypotheses, each discussed during a chapter of this document. This will lead us to the development of a specific simulation methodology taking into account the characteristics of seasonal influenza and also to a concrete new proposal to analyse large influenza VE trials based on predictive intervals. Throughout the thesis, we will illustrate our finding and apply our proposed methodologies to a large influenza VE clinical trials which will be presented at the end of this introductory chapter.

1.3 Outline

Chapter 2 of this thesis is an introduction about influenza and the development of a seasonal vaccine. It includes a review of the literature regarding the state-of-the-art in designing and analysing clinical trials (CT) data for assessing VE against seasonal influenza. In this chapter, we identify and discuss various sources of difficulties and potential causes of failure for VE CT.

Based on these findings, we explore in the next chapters four specific hypotheses as why trials could fail to show significant VE, namely:

- Are the right products selected for phase III vaccine development? (Chapter 3)

- Are the objective of VE trials too optimistic in an heterogeneous context? (Chapter 4)
- Are the classical regression models adapted to the specificities of seasonal influenza? (Chapter 5)
- Are we addressing the right question in phase III VE trials? (Chapter 6)

In Chapter 3 we discuss the influenza vaccine development process and more particularly the selection of the products to be tested in phase III. The capacity of a vaccine to produce an immune response and more particularly the post-vaccination levels of IgG antibody to haemagglutinin (HI titres) are used as a correlate of protection (COP) for VE, i.e. an indicator of the person protection against the targeted virus. We want to test whether this particular immune response is a good indicator of VE. To do so, we did an a posteriori analysis of a pooling of four efficacy trials in which both HI titres and disease occurrence data were collected for a proportion of the participants. We faced difficulties due to the differences between the trial designs and populations and the lack of exposure information. We concluded that while post vaccination HI titres were linked to VE, they were not sufficient to predict disease occurrence.

When designing a new VE CT, the required sample size is computed to reach a given power to detect a given target VE, while controlling for the type I error rate. In case the real VE is actually lower than the target level, the real power of the trial will actually be (much) smaller than expected, which will lead to an "under-powered" and often inconclusive trial. Determining, a priori, an adequate target VE is therefore a crucial step in designing a CT. This target VE is often defined for an expected number of influenza cases and will therefore depend on the attack rate (AR). A first issue is that the AR can vary from one season to the next and from region to region. An other challenge is that influenza is a multi-strain virus (Green et al. 1982). Type A virus sub-strains take the form of HxNx, for example H1N1 and H3N2. Type B viruses are divided into two lineages: Yamagata and Victoria. Genetic mutations are common and a constant monitoring of the changes in the circulating viruses is necessary (Hay et al. 2001). Typically, a seasonal influenza vaccine contains three strains, two A and one B, identified by the WHO as the most likely to circulate in the subsequent season. When the predictions of WHO are incorrect, a mismatch occurs. In Chapter 4, we propose to formalize these potential sources of heterogeneity. Based on this model, we develop a simulation algorithm which allows to study the impact of complex combinations of protocol hypotheses on the power of the trial. As we will demonstrate, by re-simulating various scenarios of a large failed phase III trial, such a tool can be very helpful in the design of new VE CT.

Classically, VE is estimated through logistic, Poisson or Cox regression models (Halloran et al. 1997). These models assume that the observations are independent and

identically distributed, given the covariate(s) if any. The subject heterogeneity described in the context of a VE trial in chapters 2 and 4 makes us question the validity of such assumption and about its impact on the VE assessment. In Chapter 5, we explore the adequacy of these classical regression models for estimating seasonal influenza VE. We investigate, first analytically then through a large simulation study, the impact of omitting these sources of heterogeneity. Interestingly, we show that when the data is highly censored, models omitting sources of heterogeneity can sometimes give better estimates than more complex models.

In chapter 6, we question whether the confidence intervals (CI) around the global VE estimates are informative enough from the public health perspective. We propose a model taking into account VE heterogeneity between geographical regions and flu seasons. Based on our model, we suggest the use of predictive intervals instead of confidence intervals in order to better reflect information about future VE. We study the properties of this new methodology via a large simulation study and apply it on data from an existing large CT.

In the last chapter of the thesis, we summarize and discuss our findings.

1.4 The Influence 65 trial

Throughout this thesis a large VE trial will be used as an illustrative example and we will apply our methodological work to this specific data set. Data from the Influence 65 trial were provided to u by GSK Biologicals. Results of this trial are published in McElhaney et al. (2013).

The Influence 65 trial was a randomized, observer-blinded study of the relative efficacy of an adjuvanted new trivalent vaccine versus the standard non-adjuvanted trivalent vaccine (McElhaney et al. 2013). The primary objective was to demonstrate the superiority of the new vaccine over the standard vaccine in preventing laboratory confirmed cases of influenza A and B in elderly adults aged $\geq$65 years. The study included 43695 subjects from 15 countries vaccinated during the 2008-2009 (year 1) and 2009-2010 (year 2) seasons. Participants were scheduled to receive one vaccine in each year, and remained in the same group in years 1 and 2. In both years, the influenza A antigens included in the vaccines were H1N1 and H3N2 strains. In year 1, the influenza Yamagata lineage B strain was foreseen by the WHO and for year 2 the Victoria lineage B strain.

Additionally, in year 1, pre and post vaccination HI titers were measured in an immunogenicity subset including 2422 and 2408 subjects in the adjuvanted vaccine and the standard vaccine groups, respectively.

In year 1, 274 and 310 subjects were infected in the new and standard vaccine groups respectively. Mean VE of 12.11% (95% CI: -3.40 to 25.29) was estimated through

a Cox regression model. Superiority of the adjuvanted could not established as the lower limit of the 95% CI for relative efficacy did not meet the predefined superiority criterion.

Unfortunately, the circulation of the pandemic influenza A H1N1 strain started during year 1 in some countries and generalized to all countries in year 2. Because of the large antigenic distance between the seasonal and the pandemic strains it was decided to exclude the pandemic flu cases from the efficacy analyses. The generalized circulation of the pandemic strain during year 2 resulted in low circulation of seasonal strains, which led to negligible AR. As a result, VE computations for the second year could not be performed.

Chapter 2

Assessing vaccine efficacy in influenza clinical trials: state-of-the-art, challenges and difficulties

This chapter is mainly based on the paper "Assessing Vaccine Efficacy in Influenza Clinical Trials: Challenges and Difficulties" By Dewé W., Benoit A and Legrand C and published in Expert Review Pharmacoeconomics & Outcomes Research, 2013 Jul;6(4):403-11.

The efficacy assessment of an investigational influenza vaccine often requires conducting large and expensive clinical trials. Particularities of influenza make singular such an evaluation and increase the complexity of the study designs and the analysis of the efficacy endpoints. Among others, these specificities are low attack rate, seasonality, multiplicity of the flu viruses, potential mutations, heterogeneity of the virus circulation in different region of the world and prediction of the vaccine composition.
In this chapter, we give an overview of the state-of-the-art for designing and analysing phase III vaccine efficacy trials data. We discuss different particularities of the seasonal influenza context and how they may impact the design, the conduct and the analysis of a vaccine efficacy trial and explains why it could fail whatever the true level of vaccine efficacy.

2.1 Introduction

Clinical efficacy of a new seasonal vaccine is studied in a large phase III randomized multi-centered trial. As a starting point for this chapter, over 100 trials were reviewed to analyse the state-of-the-art in VE phase III trials. This non-exhaustive list of trials includes observational trials as well as randomized clinical trials.

This chapter is divided into four sections. After a short introduction to VE designing, data structure and data analyses, the classical methods of analyses for VE trials are reviewed, detailed and discussed. The next section includes a non-exhaustive list of issues linked to the influenza context and how they affect the quality of the VE assessment.

2.1.1 Vaccine efficacy

In VE trials the goal is to show that the experimental vaccine reduces the risk of the infection compared to a reference. The reference can either be a placebo or a vaccine already approved for the studied virus. VE is defined as one minus the relative risk (RR) for the studied endpoint between the experimental and the comparator subjects (2.1.1) (Halloran et al. 2010). It represents the prevented proportion of infections in the subjects who received the experimental vaccine compared to those who received the reference.

$$VE = 1 - RR \tag{2.1.1}$$

Vaccines affect a population in many ways (Halloran et al. 1997). Vaccine effect for susceptibility refers to how protective vaccination is against infection. Vaccine efficacy for progression is sometimes also assessed in clinical trials. It measures how the progression of a disease is affected by the experimental vaccine. Important factors include any effect on the length of illness, a reduction in complication rate or severity of the infection. Indirect effects of vaccination are sometimes studied, although not in phase III trials. Herd immunity (Fine 1993; John and Samuel 2000), for example, refers to how a community is indirectly protected by the vaccination of some of its members by reducing the circulation of the virus. Finally, effectiveness, which measures how well a treatment works in practice, as opposed to efficacy, which measures how well it works in clinical trials, can be estimated based on surveillance data. Here, only vaccine efficacy for susceptibility will be discussed as it is the primary endpoint of phase III efficacy trials.

Success criteria for a trial can either be the non-inferiority or the superiority of the tested vaccine relative to the comparator. Non-inferiority trials are intended to show

that the effect of the new vaccine is not worse than that of an active control by more than a small pre-specified margin while offering other advantages such as a better safety profile. Superiority trials are intended to show a difference greater than a specified minimum bound. In superiority vaccine efficacy trials, success typically requires showing efficacy greater than a pre-specified clinically meaningful threshold, substantially greater than zero (Nauta 2010). The VE parameter is defined as one minus the ratio of the risks between the two groups, RR. A (one-sided) hypothesis test can be written equivalently in term of VE or RR (Dewé et al. 2013).

$$\begin{cases} H_0: & VE \leq \nu_0 \\ H_A: & VE > \nu_0 \end{cases} \iff \begin{cases} H_0: & RR \geq 1 - RR_0 \\ H_A: & RR < 1 - RR_0 \end{cases} \tag{2.1.2}$$

where $\nu_0 = (1 - RR_0)$ is the clinically-relevant threshold of success. When the test is performed at the α level of significance, efficacy is statistically demonstrated when the lower limit of the two-sided $(1-\alpha)100\%$ confidence interval of the estimated VE is larger than ν_0.

2.1.2 Design of VE trials

Phase III clinical trials often include large numbers of subjects. To achieve the recruiting rate, especially for a seasonal disease, most efficacy trials are done across multiple centres in multiple countries.

Usually, trials are run over a single season. However, in some cases, trials for influenza vaccines which continue for two to four years have been found. Various randomization schemes were found in the reviewed trials including more than one flu seasons. Vaccine can be given based on a unique randomization list. In this case, the subjects are allocated to a vaccine group at the beginning of the trial and keep the same vaccine for all the studied seasons (Edwards et al. 1994). This design might be considered as unethical as some subjects will never benefit from the vaccine. On the opposite, a new randomization list can be generated at each new season. The vaccine received can thus change every year. This design is more ethical but may create issues when vaccine history is an important covariate. Finally, some studies are particularly interested in the vaccine effect in naive subjects. In this case, subject who have previously been vaccinated are allocated to the vaccination group while naive participants are randomized between vaccine and placebo group (Keitel et al. 1997). New subjects are recruited every year and placebo subjects from the year before are added to the vaccinated group. VE can be estimated overall or for each season.

Other types of trial designs have been proposed to collect information on virus exposure. In a challenge trial, after vaccination, subjects are inoculated with the virus

and isolated in a quarantine health unit (Clements et al. 1984). The apparition and severity of the disease is closely monitored. VE estimated in this design is conditional to the knowledge of exposure. However, such trials are not allowed anymore in the context of influenza as the disease can cause severe complications. In past studies, the population that could be included in a challenge trial trial also very restricted: they could never be performed in the populations most at risk from influenza, the elderly and young children. In household contacts studies, disease occurrence in the family members of index cases are collected, allowing for partial information about disease exposure(Longini et al. 1982; Esposito et al. 2003). In Chapter 3 the importance of some knowledge about exposure will be discussed.

In the trials discussed here, the censoring mechanism is primarily administrative: subjects who did not have the event by the end of the trial are considered as censored. The cumulative incidence of influenza infection by the end of the season, i.e. the attack rate (AR), is usually low resulting in highly censored data. This point will be of particular interest in Chapter 5. Withdrawal from the study or lost to follow up constitute the second censoring mechanism. Both these types of censoring are usually considered as non-informative. A third mechanism of censoring, departure from the trial due to a severe adverse event might be informative. Indeed, events such as hospitalization, pneumonia, heart attack or death might be related to an episode of influenza. In a healthy adults population, they are rare and can be ignored while in elderly, the occurrence of such events is often analysed separately (Gross et al. 1995; Rothberg et al. 2008).

2.1.3 Limitations of our work

Due to the characteristics of influenza, there are limitations to the modifications we can bring to the design of VE trials. The first limitation is ethical. While inoculating the virus to all trial participants would insure that all of them have been exposed, such a procedure is consider as non-ethical and not authorized by the health regulatory authorities. Also, in populations where annual vaccination is recommended, such as the elderly, the use of a placebo vaccine is not allowed and the reference vaccine is then the standard of care. This particular topic will be discussed in Chapters 3 and 4.

A second limitation is economical and practical. Because seasonal influenza only affects a small portion of the population, VE trials usually include many subjects (over 43000 for trial Influence 65, see section 1.4 in the previous chapter). Because of the large trial sizes, collecting a lot of information about trial participants, or taking blood samples in all of them is very costly and not feasible in practical.

A third limitation is linked to the vaccine effect duration. The impact of previous vaccinations on actual vaccine efficacy is controversial. This topic will be discussed in Section 2.2.4 of this chapter. As a result, alternating between different vaccines,

such as in a cross-over design, is not feasible and parallel designs are preferred. We do not rethink the two-arms randomized double-blind parallel design generally used for phase III VE trials.

Finally, the motivation for our researches is the failure of several phase III VE trials. We mainly concentrate our researches on this phase of development.

2.2 Statistical models to analyse vaccine efficacy data

In this section we propose an overview of the statistical models classically used to analyse VE data. The assumptions of those models as well as their applicability to the specific case of seasonal influenza will be discussed and explored in Chapter 5.

The risks ratio between the experimental and the comparator vaccine groups, RR, is $\frac{\psi_1}{\psi_0}$. Depending on the level of information collected about the occurrence of the disease, ψ_0 and ψ_1 can be modelled differently. When only the status (infected or not) of the participants at the end of the trial are known, ψ_0 and ψ_1 are respectively the cumulative incidence in the reference and the comparator groups. If time-to-event data are collected, ψ_0 and ψ_1 are defined as incidences or hazard rates.

2.2.1 RR as a ratio of risks: Cumulative incidence

Estimating the relative risk of infection based on the cumulative incidence requires only information about the number of subjects and influenza cases in each group during a fixed surveillance period. The cumulative incidence is calculated as the ratio between the number of influenza cases occurring during the surveillance period and the number of subjects in each group. VE is defined as one minus the cumulative incidences ratio between the experimental and the comparator vaccines.

Non parametric approaches

In the early seventies, VE was computed from the observed attack rates ratio of the experimental and comparator groups (Hobson et al. 1973). A 2 x 2 contingency table was generated from the data. To assess the significativity of the result, a Pearson's chi-square test was done, with the null hypothesis being a relative efficacy of one. This simple approach presents several drawbacks.

The first drawback of the chi-square test is that it only allows for the null hypothesis of equal efficacy. Yet, evidence of vaccine efficacy often requires a relative risk significantly greater than 1 between the tested vaccine and the comparator (Chan et al.

2003). Confidence interval can be built based on the asymptotic normality of the log of the ratio of two proportions (Chan et al. 2003). Koopman (1984) proposed a method for approximating confidence intervals for the ratio of two binomial proportions based on two independent binomially distributed random variables. The confidence limits, built from the chi-square distribution quantiles, cannot be calculated directly from a formula, but must be derived iteratively. The Koopman's chi-square method has been used in several influenza vaccine efficacy trials in the nineties and is implemented in several statistical softwares (Belshe et al. 1998, 2000). More recently, the exploding power of computers has made possible the return to exact methods. Exact confidence intervals conditionally to the total number of infection cases are now commonly reported in vaccine efficacy trials (Ohmit et al. 2006). An exact confidence interval is constructed by inverting the critical region of Fisher's exact test (Ewell 1996).

The second drawback of the chi-square test is that it does not take into account the possible presence of nuisance factors. The Mantel-Haenszel method allows the calculation of a weighted statistics adjusted for the effect of a stratification factor (Landis et al. 1998). It estimates the association between the vaccine group and the event variables, adjusted for the effect of the stratification factor. A common vaccine effect can be estimated across the confounder strata using a weighted mean of a measure of association. This test has been used in a few trials for the last two decades (Kawai et al. 2003).

The biggest weaknesses of these methods is that they only allow the stratification over categorical variables and do not allow the estimation of the covariates effects.

Parametric approaches

The logistic regression models are the most commonly reported methods in the analysis of influenza VE data (Edwards et al. 1994). They rely on cumulative incidence and as such do not take into account the time of exposure. The estimated odds ratio (OR) is often interpreted as an estimate of the RR despite their differences. However, when the outcome risk is small, such as for seasonal influenza, the difference between the OR and the RR is negligible and they can be interpreted similarly (Zhang and Kai 1998; Lachin 2011). Figure 2.1 shows the evolution of the OR in function of the prevalence of the event in the comparator group, and for different values of the RR. Estimated RR can be derived from the OR (Zhang and Kai 1998) from:

$$RR = \frac{OR}{(1 - AR_0) + (AR_0 \times OR)} \tag{2.2.1}$$

where AR_0 represents the probability of event in the comparator group.

As the RR between the tested vaccine and the comparator is expected to be smaller than 1 (reduction of the risk with the new vaccine), the bias implies underestimation of RR, i.e. an overestimation of VE. This particular effect will be examined in Chapter 5.

The logistic regression model is a special case of the generalized linear models, in which the the link function is the logit link, $\log(\frac{\pi}{1-\pi})$, where π is the proportion of events (Lachin 2011). Estimation of the model parameters is done through maximum likelihood methods (Agresti 1996).

Based on the following model,

$$\text{logit}(\pi|X_1 = x_1, X = x) = \beta_0 + \beta_1 x_1 + \beta' x \tag{2.2.2}$$

where β_0 is the baseline risk parameter, X_1 is the vaccine group factor, β_1 the regression parameter for the vaccine effect, X contains the values and β the regression parameters, β_2 to β_{q+1}, for the other q covariates. VE is defined as $1 - OR_1 = 1 - \exp\beta_1$.

The center effect can be included in different ways in this model (Localio et al. 2001). A fixed-effect logistic regression includes the center as an additive factor. As such, centres included in the trial represent themselves rather than a sample of centres from the population of all centres. We consider the case of J centres with $j = 1, ...J$. The fixed-effects logistic regression model is of the form:

$$\text{logit}(\pi_j|X_1 = x_1, X = x, J = j) = \beta_{0,j} + \beta_1 x_1 + \beta' x \tag{2.2.3}$$

where $\beta_{0,j}$ is the baseline risk in the j^{th} center, $j = 1, ...J$.

In a conditional logistic regression, center effects are considered as nuisance effects and not estimated. This method is only appropriate when there are many subjects per center as those where no event are observed do not contribute to the likelihood. A mixed effect logistic regression assumes that the centres included in the trial are a random sample from the population of centres. An overall vaccine effect is estimated as well as a random center-wise baseline risk.

$$\text{logit}(\pi_j|X_1 = x_1, X = x, J = j) = \beta_0 + \beta_{0,j} + \beta_1 x_1 + \beta' x \tag{2.2.4}$$

where β_0 is the mean baseline effect and $\beta_{0,j}$ are the random center effects assumed to be normally distributed with mean 0 and variance $\sigma^2_{\beta_0}$. Interaction fixed or random effects between centres and other covariates can be included in all models.

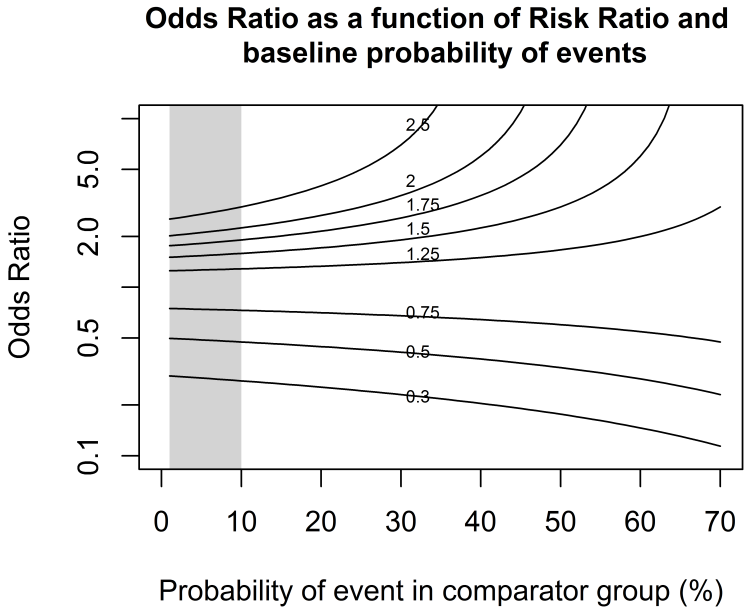

Figure 2.1: Relationship between OR and probability of event in the reference group for varying levels of RR. The shaded band represents the expected AR for seasonal influenza.

2.2.2 RR as a ratio of incidence rates: Person-time approach

The incidence rate (IR) is the risk of experiencing an infection during a given time unit. As opposed to relying on cumulative incidences, counting the number of events by time-unit allows for different exposure times among the subjects. In large studies where the disease incidence is low, it is assumed that the number of events in the vaccine and control groups may be approximated by independent Poisson distributions. Conditional on the total number of events, the number of events in the vaccinated group follows a binomial distribution (Lachin 2011).

Non parametric approaches

Confidence intervals for a ratio of incidence rates can be built based on the asymptotic normality of the logarithm of such a ratio (Ewell 1996). Modern computing power allows for an exact test to be constructed conditionally on the number of events, even for large studies (Hoberman et al. 2003). As already mentioned, the non model-based method presents the major drawback of not allowing the adjustment for continuous covariates or more than a few categorical factors.

Parametric approaches

The multiplicative Poisson regression model is also a special case of the generalized linear models with the logarithm as the link function. The time of exposure, t_i, is included in the model as an offset (Bridges et al. 2000; Frome 1983).

$$\log(\mu|X_1 = x_1, X = x, T = t) = \log(t) + \beta_0 + \beta_1 x_1 + \beta' x \tag{2.2.5}$$

where μ is the rate of events, t is the time of exposure , β_0 is the baseline event rate for one unit of time, X_1 is the vaccine group factor, β_1 the regression parameter for the vaccine effect and X and β are the variables and the regression parameters for the other covariates, with $\beta = (\beta_2, ...\beta_{q+1})$. VE is defined as $1 - IR_1 = 1 - \exp\beta_1$.

The Poisson regression model allows for the inclusion of continuous and categorical covariates. Incomplete observations from drop-out subjects can be included in the model as long as the censoring is not informative. As parameterized, the risk of infection over time is assumed to be constant, conditionally on the covariates included in the model.

As for the logistic regression model, the center effect can be included as a fixed or a random effect into the Poisson regression model. Estimation methods include maxi-

mum likelihood and restricted maximum likelihood for the models including random effects.

2.2.3 RR as the ratio of two forces of Infection: Hazards ratio

A third way to look at the data is to consider the time between the start of the observation period and the day of onset of the event. Subjects who have not experienced the event at the end of the follow-up period are considered censored . Time-to-event data are usually described in terms of survival and hazard functions.

We define T as the time between the beginning of the surveillance period and influenza infection onset. The censoring time, C, is the time between the start of the surveillance period and censoring, either the last contact with the subject or the end of the surveillance period in the case of administrative censoring. The observed time-to-event Y is the minimum between T and C and the censoring indicator variable δ is set to 1 when the subject was infected by influenza during the surveillance period, 0 else.

We define $f(t)$ as the probability density function of the event time t and $F(t)$ the corresponding cumulative distribution function

$$F(t) = P(T \leq t) = \int_0^t f(u)du \tag{2.2.6}$$

The survival function, $S(t)$ gives the probability that a subjects has not yet experienced influenza infection at time t

$$S(t) = 1 - F(t) = P(T > t) \tag{2.2.7}$$

The hazard function, $h(t)$ gives the instantaneous rate at which influenza infection occurs for subjects still at risk:

$$h(t) = lim_{\triangle t \to 0} \frac{P(t \leq T < t + \triangle t | T \geq T}{\triangle t} = \frac{f(t)}{S(t)} \tag{2.2.8}$$

The cumulative hazard function, $H(t)$, gives the accumulated risk of infection by time t

$$H(t) = \int_0^t h(u)du \tag{2.2.9}$$

The impact of the covariates is often modelled via this hazard function, like in the Cox regression model (Cox 1972), one of the most popular model for time-to-event data.

VE represents the reduction in the instantaneous hazard of infection in the experimental group compared to the reference group, the hazard ratio (HR).

Non parametric approaches

Time-to-event data can be analysed through non-parametric methods. In this case, no assumptions are made on either the baseline or the covariate part of the hazard function.

One possible non parametric way to estimate $S(t)$ is the Kaplan-Meier curve(Treanor et al. 2011). The log-rank test can be used to test for evidence of a vaccine effect. Extensions of this test, using different weight repartitions over the timeline of events, are possible. However, as with all non-model based methods reviewed in this section, working with multiple categorical covariates is not possible. Beran estimator (Beran 1981) is a non-parametric asymptotic estimation of $S(t)$ which makes the inclusion of a continuous covariate possible via kernel functions. Non-parametric methodologies cannot be used either to explore (and adjust for) the effects of several variables. Finally, no parameter to summarize the magnitude of the vaccine efficacy is available with these methods.

Parametric approaches

Several modelling statistical techniques for modelling the relationship between time-to-event and several explanatory variables have been developed. These techniques includes parametric and semi-parametric methods. The parameters β are usually assumed to be time-invariant, although extensions to estimate time-dependant coefficients have been proposed. One possible choice is to assume a time-constant multiplicative relationship between the underlying hazard function and the log-linear function of the covariates, leading to hazards proportionality (PH) over time.

$$h(t|X_1 = x_1, X = x) = h_0(t) \exp(\beta_1 x_1 + \beta' x), i = 1, ..., n \tag{2.2.10}$$

where $h(t)$ is the hazard function, X_1 is the vaccine group factor, β_1 the regression parameter for the vaccine effect and and X and β are the variables and the regression parameters for the other covariates, with $\beta = (\beta_2, ...\beta_{q+1})$. VE is defined as $1 - HR_1 = 1 - \exp \beta_1$.

In the Cox regression (Cox 1972), a semi-parametric model, no assumptions are made about the nature or shape of the hazard function. The baseline hazard is unspecified and the regression coefficients are estimated by the method of the partial likelihood (Machin et al. 2006).

Fully parametric PH model have the same form as the Cox regression model but, in addition, assume a particular distribution of the time-to-event (Figure 2.2). The exponential model corresponds to assuming that the survival times are exponentially distributed leading to a constant baseline hazard. Weibull distribution of the survival times coincides with a monotone baseline hazard. In the lognormal model and the log-logistic model the baseline hazard has the value 0 at $t = 0$, increases to a maximum and then decreases as t becomes large. Completely parametric models are more efficient than Cox regression if the underlying model is correct (Nardi and Schemper 2003). However, their assumptions seem rarely tenable, especially in the case of a seasonal infection such as influenza.

In PH models, center effects can be included as strata, fixed effects or random effects (corresponding to the shared frailty model) (O'Quigley and Stare 2002). The first solution is only possible for categorical covariates but offers the advantage of not assuming PH between strata. For fixed and random effects, PH is assumed between the levels of the covariates.

The stratified model

$$h(t|X_1 = x_1, X = x, J = j) = h_0(t, j) \exp(\beta_1 x_1 + \beta' x) \tag{2.2.11}$$

where $= h_0(t, j)$ is the j^{th} center baseline hazard, $j = 1, ...J$.

The fixed effect model

$$h(t|X_1 = x_1, X = x, J = j) = h_0(t) \exp(c_j + \beta_1 x_1 + \beta' x) \tag{2.2.12}$$

where c_j is the j^{th} center effect, $j = 1, ...J$.

The shared frailty model

$$h(t|X_1 = x_1, X = x, J = j) = h_0(t) u_j \exp(\beta_1 x_1 + \beta' x) \tag{2.2.13}$$

where u_j is the random effect for the j^{th} center, $j = 1, ...J$ and the random variable U has probability density f_U, called the frailty distribution.

2.2.4 Sources of heterogeneity

Sources of heterogeneities are numerous in the context of influenza. The main ones and their effects on VE are reviewed in this section. Figure 2.3 represents the sources of heterogeneity as well as their interactions. In chapter 4, we develop a simulation methodology fitted to this schema.

In red, we present the main components of a phase III VE CT: trial subjects are vaccinated either with the comparator or the experimental vaccine. During the viruses

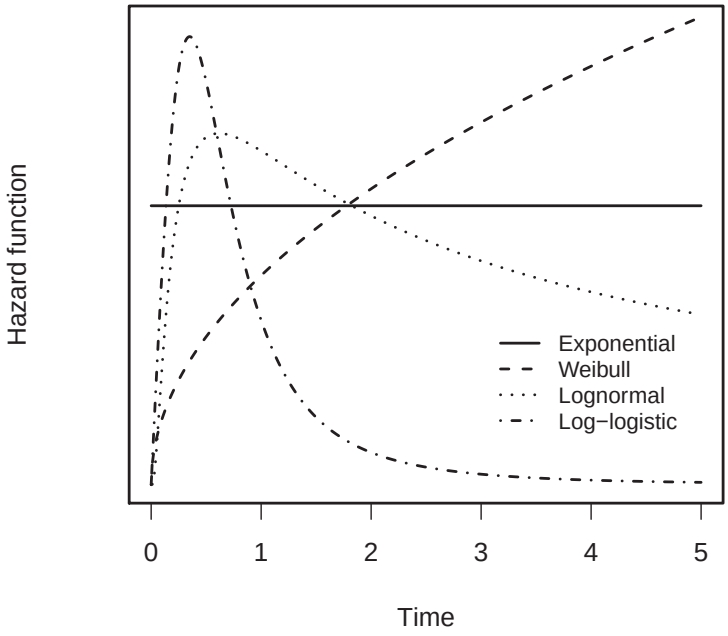

Figure 2.2: Hazard functions examples, $h(t)$, for exponential, Weibull, lognormal and log-logistic time-of-event distributions.

circulation period, subjects from both vaccination groups may get infected and suffer from clinical influenza.

The statistical analyses applied to trial data are presented in pink. VE is estimated through logistic, Poisson or Cox regression models and decision about the efficacy of the experimental vaccine is taken based on a confidence interval for this quantity. Statistical analyses models have been presented in section 2.2 of this chapter and the quality of their estimates of VE in the context of influenza will be assessed in Chapter 5.

Several sources of difficulties arise in this simple scheme and will be discussed in the subsequent sections: mismatching between the vaccine and the circulating strains, heterogeneity in the intensity of circulation and the infectiousness of the circulating viruses and heterogeneity in the vaccine efficacy mechanisms. The impact of these sources of heterogeneity on the estimation of VE will be assessed and discussed in Chapter 5.

The green part of the schema summarizes the objective of a correlate of protection study. Prior to phase III, the clinical development of a new vaccine is based on its potential to induce an immunogenicity response. This response is evaluated through a post vaccination blood sample analysis. The goal of a COP study is to study the link between the post-vaccination immunogenicity response and the risk of infection with the virus strain of interest. One source of heterogeneity in this process occurs at the blood sample analysis phase: between laboratories and inter-run variances are two potential sources of difficulties. We present a COP study for one strain of influenza in Chapter 3. This study is based on a pooling of four efficacy trials and the problem of between trials heterogeneity, due to laboratory differences for example, will be illustrated and discussed.

External sources of heterogeneities are presented in blue. The main one is the subject himself. This topic will be presented in section 2.2.4 and its impact on COP and VE estimations will be discussed respectively in Chapters 3 and 5. Other possible sources of heterogeneities are linked to the passage of time: first, we point out the seasonality of influenza, which will be of particular interest in Chapter 5. Second, between seasons and within a season, circulating strains of viruses tend to change, due to mutations for example. Impact of this effect on the estimation of VE and solutions to take them into account in the design and the analysis of VE trials will be presented respectively in Chapters 4 and 6. Finally, the effect of the vaccine itself might be impacted by the time elapsed since vaccination. For example, there is a delay between vaccination time and protection time. Also, VE may slightly decrease with and influenza season, an effect called "waning". We do not discuss these issues.

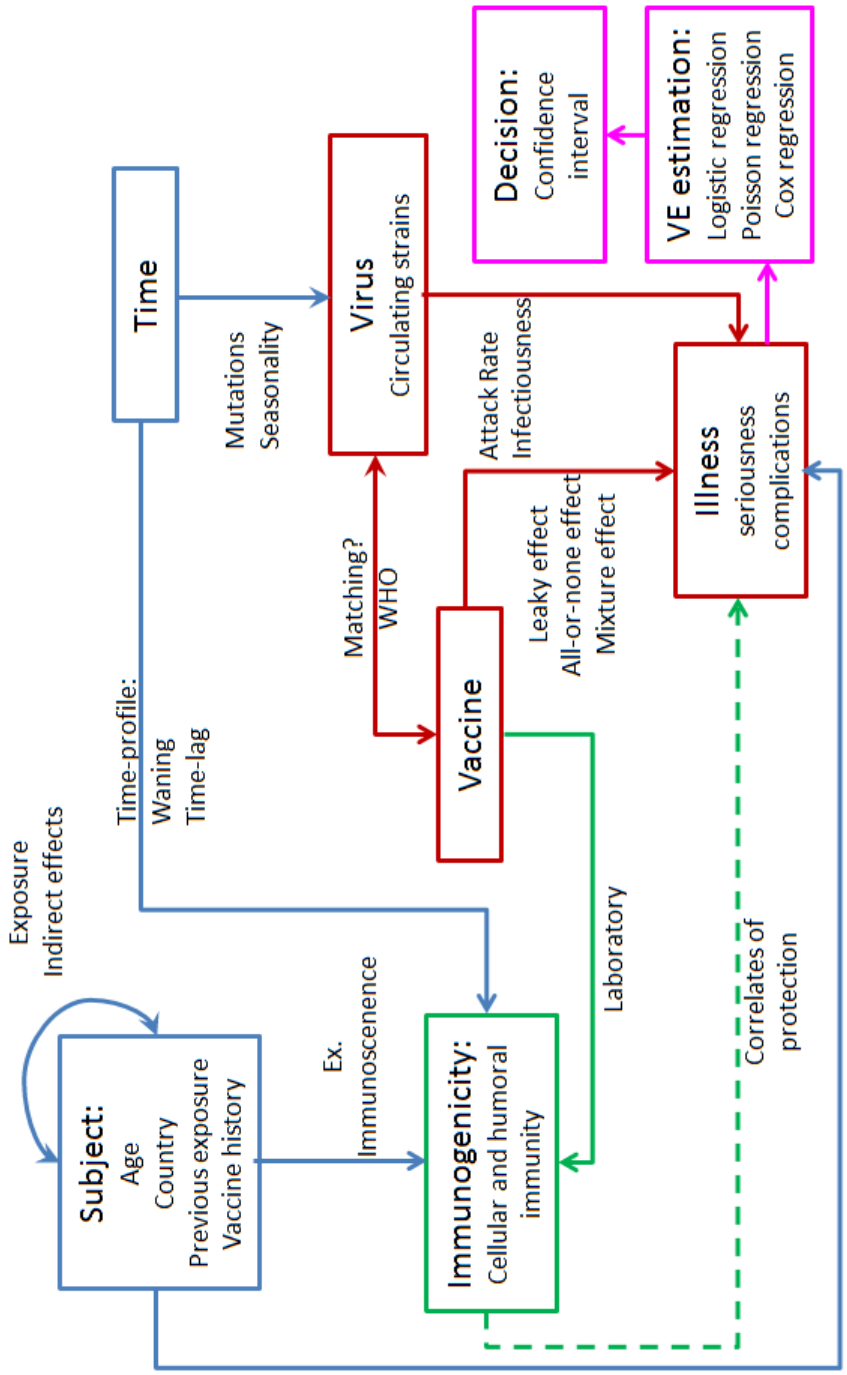

Figure 2.3: Sources of heterogeneity in seasonal influenza vaccination

Subject

Not everybody is equally affected by influenza. Individual characteristics, such as age, preexisting conditions or behaviours influence the probability of infection, the response to vaccination and the severity of illness (Figure 2.3).

Influenza infection begins with exposure of the upper respiratory tract to the virus. The infected host responds almost immediately to exposure, including through the immune system (WHO 2012). People who have weaker immune systems, such as the elderly, sick people, or young children, may be quickly overwhelmed by the virus (Weinstein et al. 2003).

Vaccine effects are also mediated by the immune system. Given the same vaccine, young children and elderly subjects appear to have a weaker antibody production than healthy adults, resulting in a weaker protection. Immunosenescence in the elderly has been studied extensively (Sambhara and McElhaney 2009; McElhaney et al. 2006). It has been suggested (Pawelec 1999) that young children might suffer from an similar condition.

Influenza infection is conditional to exposure to the virus. Flu can spread only short distances between people, by direct contact with contaminated secretions or when the virus is sneezed or coughed up with droplets of nasal secretions or saliva carried over short distances into the air (Barker et al. 2001). As a result, different behaviour patterns lead to varying exposure levels. For example, the highest infection rates are found in children from 5 to 9 years old, who are likely to be in close contact with other sick children and in institutionalized elders. On the opposite, non-institutionalized elders tend to spend more time at home and have fewer contacts with infected persons. Finally, health workers or people in contact with children are at greater risk of exposure. Whereas age is recorded in clinical trials, occupation and and number of contacts are usually not. Also, different sub-populations (age categories) tend to be investigated in different trials. As a result, direct comparison of different age groups is not possible and inter trial comparisons are flawed due to multiple disparities between trials. This specific point will be discussed in Chapter 3.

Virus

Heterogeneity also arises from virus heterogeneity (Figure 2.3). Several types and sub-types of influenza virus coexist and continually evolve. Mutations (antigenic drift) and combination between viruses (shift or exchange of a genomic segment) modify the structure of the virus surface antigens (Green et al. 1982). Current vaccines contain three strains (A/H1N1, A/H2N3 and B) recommended annually by WHO. The selected strains are the most likely to circulate in the subsequent season. Due to the high mutation rate of the virus, a particular vaccine formulation looses efficacy rapidly and annual vaccination is recommended. (Carrat and Flahault 2007).

The level of infectiousness and the disease severity vary from one influenza strain to another. For example, type A viruses are the principal causes of larger epidemics while the type B viruses are often limited to localized outbreaks (WHO 2012). The A strains are also more prone to mutations than B strains. Because of the high mutation rate of the virus, infection occurring at the beginning of the season might not be caused by the exact same virus as later in the season. Also, the repartition of the circulating strains differs by geographical regions (Figure 2.4). For those reasons, heterogeneity in VE between country is likely and should be taken into account in analysing international trials. This particular point will be discussed in Chapter 6.

Vaccine

How vaccines protect against influenza is not completely understood. Smith et al. (1984) consider two mechanisms of protection, Type I (*leaky*) and Type II (*all-or-none*). In the former, vaccination is assumed to reduce the instantaneous disease rate in all vaccinees by a constant proportion. In the latter, vaccination is assumed to provide a constant proportion of individuals with complete immunity.

The mechanism of vaccine protection has an impact on the choice of efficacy measures (Halloran et al. 1997). On one hand, with a *leaky* vaccine, everybody has the same baseline risk of infection. Vaccination multiplicatively decreases the risk in all vaccinated subjects. As the season advances, the level of exposure grows and as a result, cumulative infections rates increase (Figure 2.5). So, assuming a *leaky* effect, the most appropriate models should take into account the time of exposure, such as the Poisson and Cox regression models. On the other hand, an *all-or-none* vaccine does not confer the same protection to all individuals. While individuals are completely protected, others do not get any benefit. As a result, as the season advances and exposure increases, the susceptibility of the vaccinated group will decrease possibly all the way down to zero, once all non protected subjects have been infected. In presence of a *all-or-none* effect, models assuming PH are not appropriate (Figure 2.5).

Most likely, influenza vaccination results in a mixture of the two mechanisms, conferring no protection, heterogeneous partial protection and full protection to varying proportions of the subjects. Natural immuno-protection (due to previous exposure to a similar virus for example) probably leads to a similar pattern of protection heterogeneity in the non-vaccinated subjects. Halloran et al. (1997) suggests the use of survival models including a degenerated frailty with a point mass at 0 to account for vaccine protection mechanisms. This model will be discussed in Chapter 5.

Vaccine effect could also decrease along the flu season. This effect is called "waning" and contradictory results exist on the topic: Declining levels of antibodies in the months following vaccination have been reported (Belongia et al. 2014). A review

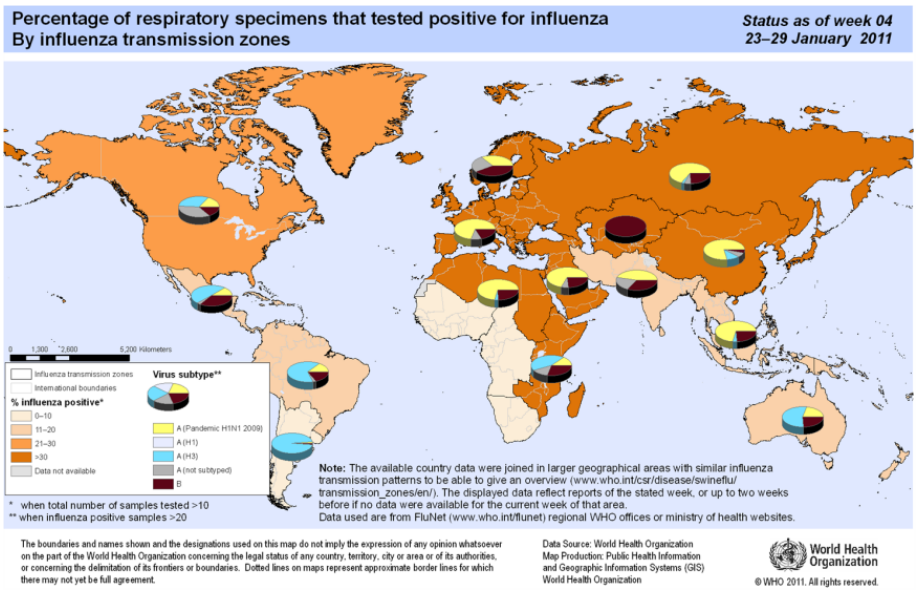

Figure 2.4: Percentages of respiratory specimens that tested positive for influenza by strain (pie charts) and by influenza transmission zones between January 23 and 29, 2011.

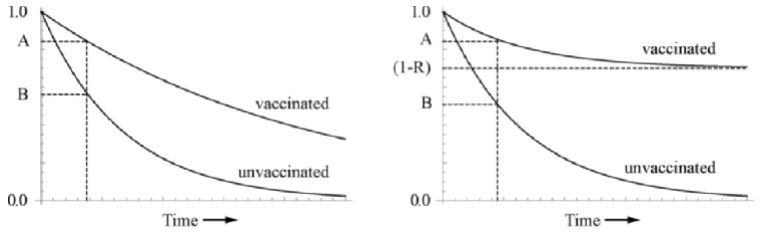

Figure 2.5: Proportion of individuals without disease by time since start of the surveillance period under the leaky (left) and the all-or-none (right) mode of action. From Halloran et al. (2010).

papers by Skowronski et al. (2008) reports, on the contrary, no such effect. We will see in Chapter 5 that time varying covariate effect may be difficult to detect as it can be confounded with latent individual heterogeneity.

Context

Context variables also induce heterogeneity in the effectiveness of an influenza vaccine. Vaccination recommendations differ between countries. For example, in the US, influenza vaccination is recommended for all persons aged 6 months and older while in some European countries, such as in Belgium, only the elderly are recommended to be immunized. As a result, vaccine coverage is higher in the US, resulting in decreased virus circulation and indirect protection of the unvaccinated subjects.

The impact of vaccine history on the actual vaccination effect has been frequently discussed in the influenza research community. Annual vaccination is recommended in the populations at risk of complications. While it has been suggested that protection after annual influenza vaccination would successively decrease (Hoskins et al. 1979), a review study by Beyer et al. (1999) did not confirm such results. If vaccine history affects present vaccine efficacy, difference in efficacy could be observed in countries making different recommendation policies.

2.2.5 Presence of uncontrollable nuisance sources

Vaccine efficacy estimation is also difficult because of nuisance factors and unknown information. Non-matching between vaccine and circulating strains, the lack of a surrogate of protection, seasonality of the epidemic and low attack rates are factors discussed in this section.

Non-matching between vaccine and circulating strains

As mentioned in section 2.2.4, WHO recommends vaccine strains based on predictions that are not always correct and mismatching can occur between circulating and vaccine strains. If the antigenic distance between them is large, the vaccine will likely offer no protection to the vaccinee. Thus, the effectiveness of the annual influenza vaccine varies from year to year due to changes in the circulating influenza strains. Current measures of antigenic distance are based on ferret antisera hemagglutinin inhibition assays (Gupta et al. 2006). If the distance is too large, influenza case is considered as non-matching and, depending on the protocol, might not be taken into account in the efficacy estimation. Pharmaceutical companies are currently developing novel influenza vaccines that would protect against more virus strains, either by including

more antigens in their vaccines or by proposing formulations that would confer cross-protection. However, as we will see in Chapters 3 and 4, the development of such a product is very challenging.

No surrogate of protection

Immunity to influenza infection is a multi-factorial phenomenon. Specific IgG antibodies, cell-mediated immunity and local antibodies all seem to play a role in the protection against influenza (Nauta 2010). Unfortunately, the exact role of each contributor is not yet fully understood. The level of IgG antibody estimated by the haemagglutination inhibition titer (HI) is the easiest to measure and is usually the primary endpoint in immunogenicity trials. Seroconversion, seroprotection and mean fold increase (pre to post vaccination) are generally accepted as indicators of VE. For urgently needed vaccine (pandemic vaccine for example) and for annual registration, such an information is sufficient for the product to be approved for licensure. However, recent vaccines designed to increase the antibody responses have failed to show superior efficacy with respect to the standard of care in large phase III trials (McElhaney et al. 2013). This matter will be discussed in Chapter 3.

Seasonality

In temperate climates, influenza infections are characterized by seasonality. The disease is marked seasonal peaks, typically during the winter months (Lofgren et al. 2007). The risk of infection is thus time-dependent. Statistical methods can be affected by this characteristic. This issue will be discussed in Chapters 4 and 5.

Low attack rates

VE clinical pivotal trials are characterized by large sample sizes (over 43000 subjects for trial Influence 65 for example), to reach a targeted statistical power to detect the outcome of interest. In the context of influenza, attack rates in the unvaccinated population may vary widely and is not easy to predict accurately. Also, for the high risk population, the use of placebo may be considered as non-ethical, especially in the countries where vaccine policies recommend vaccination for those groups. All subjects in the study are therefore vaccinated either with the tested vaccine or an already marketed comparative vaccine, resulting in even lower infection rates. As a result, the observed attack rates can be low and the sample sizes required to gather enough events, very large. Observations are thus numerous, but only a small fraction of them are informative. So, to reach the desired sample size before the start of the influenza season, subjects are often recruited from multiple study centers.

In the case of very efficacious vaccines, it may be that no or very few infections occur in the experimental group. In this case, there is a (quasi)complete separation of response and non-responses by the vaccine variable. Heinze and Schemper (2001, 2002) show that this may lead to estimation issues due to the monotony of the likelihood function. However, this situation is very rare, especially in influenza infections, since more characteristic of small samples (n<200).

2.3 Conclusion

In this chapter, we presented information about seasonal influenza epidemics and the methodology to assess VE against this infection. We showed that influenza presented many characteristics that render the design and analysis of CT challenging. In the subsequent chapters, specific difficulties will be detailed and further explored. The impact of the contradiction between the heterogeneity of the seasonal influenza context and the quite simple models used to analyse influenza VE trial will be studied in Chapter 5. In chapter 6, we present a new methodology that accounts for some sources of heterogeneity in the design, the analysis and the decision-making stages of a phase III VE trial.

Summary of Chapter 2

- Influenza is a respiratory disease due to a viral infection against which vaccination against influenza is recommended, especially for high risk populations (elderly, children, etc.)
- VE is studied in large multi-centric international phase III trials and data are usually analysed through logistic, Poisson or Cox regression models
- The context of seasonal influenza is heterogeneous and this heterogeneity may be the cause of past failed trials

Chapter 3

Correlate of protection for seasonal influenza

This chapter is based on the paper "Haemagglutination inhibition antibody titres as a correlated of protection against seasonal A/H3N2 influenza infection" by Benoit et al., and submitted to Open Forum Infectious Diseases

New influenza vaccines are classically developed based on their capacity to induce an immunogenicity response. However, recent trials failures have emphased the lack of evidence relating serum hemagglutinin-inhibition antibody levels to the risk of influenza disease.
In this chapter, we present a a correlates of protection analysis based on pooled data from four randomized trials. We develop a logistic model to evaluate the titre of A/H3N2 HI antibodies as a correlate of the occurrence of A/H3N2 disease. We then build a receiver operating characteristic curve to identify a potential cut-off titre between protection and no protection.
We confirm that there is a relationship between the occurrence of A/H3N2 disease and HI antibody responses. Including age and epidemic intensity as covariates, a four-fold increase in titre is associated with a two-fold decrease in the risk of A/H3N2 disease. We find that an universal threshold of protection does not seem realistic since the protection level depends upon other factors, such as subjects characteristics and level of exposure.

3.1 Introduction

The HI titres estimating IgG antibody levels are traditionally used as markers of efficacy in the earlier phases of influenza vaccine development. Seroconversion and mean fold increase in HI titres(pre vs. post vaccination) are generally accepted as predictive of vaccine efficacy. HI titres are considered a correlate of protection (COP), but the mechanism of protection is not fully captured by this measure and the predictive power of HI titres has not be proven in clinical trials (Qin et al. 2007). However, pharmaceutical companies strongly rely on this measure in order to develop their products.

Every season, virus strains selected to be included in the annual vaccine may change, making it impossible to perform a study to demonstrate the efficacy of the new vaccine in the following season. Therefore, each year, vaccine are licensed based on immunogenicity measures that assume post vaccine HI titres above a defined threshold will be sufficient to prevent influenza. Criteria recognized by the health European and US regulatory authorities are presented in table 3.1.

	EMA (Europe)		**FDA (US)**	
	One criterion out of the 3		**Both criteria**	
	Point estimates		**Lower limit of the 95% CI**	
	18-60y	**≥60y**	**18-65y**	**≥65y**
Seroconversion factor [a]	>2.5	>2	NA	NA
Seroconversion rate [b]	>40%	>30%	>40%	>30%
Seroprotection rate [c]	>70%	>60%	>70%	>60%

Table 3.1: FDA and EMA criteria for a successful phase II trial. [a] Geometric mean titre increase ratio. [b] percentages of subjects with at least a factor 4 between pre and post vaccination HI titres. [c] percentages of subjects with HI≥40.

Assuming a relationship between HI titres and VE, pharmaceutical companies are attempting to develop new generations of flu vaccine or to extend licensures, based on the sole immune responses. Recently however, the superiority of a new vaccine candidate over a non-adjuvanted vaccine based on higher HI responses did not result in superior vaccine efficacy in the large phase III trial influence 65 (McElhaney et al. 2013). As a result, further investigation is needed to determine whether HI responses are a reliable endpoint to optimize the formulation of new vaccines and also to market vaccines whenever performing a VE study is not possible.

Ideally, COP should be studied in a well-designed challenge trial as follow: first subjects are vaccinated either with the experimental vaccine or the reference vaccine; seconds, 21 days later, their immunogenicity response is measured through a blood sample; third, they are inoculated with the virus strain(s) of interest and fourth they are observed to determine the occurrence of the disease. However, such studies are believed to be unethical and no longer approved. Therefore, COP are studied as an ex-

ploratory endpoint of phase III trials: immunogenicity analyses are performed based on the fraction of the study participants on whom both pre and post vaccination HI titres are measured as well as the primary efficacy endpoint.

The main issue with COP as an exploratory endpoint of efficacy trials is that little information is collected on the study participants. Influenza infection is conditional on exposure to the virus. Contrarily to a challenge trial, efficacy trials provide no information about the level of exposure of a subject. We know that infected subjects were exposed to influenza. However, non-infected subjects are a mixed between two populations: exposed subjects protected by the vaccine and non-exposed subjects who were either protected or not.

Qin et al. (2007) provide a framework to assess correlates in vaccine trials. In their classification, while a COP should be statistically correlated with the endpoint of interest, a surrogate of protection (SOP) has a predictive power for the endpoint of interest and is independent of the setting. Thus, to validate any correlate as a predictor of efficacy it would be desirable to pool databases of different studies conducted in different settings (Gilbert et al. 2008). The downside of pooling data from several efficacy trials ran in different sub-populations and flu seasons, with different vaccines, is that it leads to interpretation difficulties since an observed effect could be attributed either to a covariate of interest or to inter-trial differences.

Furthermore, pharmaceutical companies would like to identify a threshold of protection titre above which a person would be considered as protected. An HI antibody threshold of 1:40 is generally recognized as corresponding to a 50% reduction in the risk of influenza, based on a challenge study in adults conducted by Hobson et al. (1972). However, there is no consensus on the definition of "protection" with some studies defining protection as a pre-defined risk reduction (usually 50%) and some studies defining protection as the titre level providing the best separation between influenza cases and non cases (Dunning 2006). To date, there is little agreement on what represents "protection" or any cut-off antibody titre between protection and non-protection based on contemporary data from influenza vaccine trials.

Here we describe a COP analysis of pooled data from four randomized trials of seasonal influenza including 7730 subjects. Our analysis is focussed on strain A/H3N2. We model the data through a logistic regression model. To take into account the exposure to the virus, we include season strength as a covariate. Based on our model we build a receiving operating characteristics (ROC) curve in order to assess whether post vaccination titres are associated with the prevention of the next influenza episode (Zweig and Campbell 1993; Zou et al. 2007). We attempt to derive a post vaccination HI titres threshold between protected and non-protected subjects. We discuss our results and provide recommendation for running future COP trials.

3.1.1 Materials and Methods

The analysis was based on four phase III trials, two in subjects aged 18-64 years, one in subjects aged 18-49 years, and one in subjects aged $\geq$ 65 years. In each efficacy study, immunogenicity assessments were performed on a randomly assigned sub-cohort, and our analysis was performed on the per-protocol sub-cohort (including subjects who met eligibility criteria, complied with the protocol, received any dose of either vaccine, and for whom data were available):

1. Beran et al. (2009a) performed a randomized, double-blind, placebo-controlled study of the efficacy of trivalent vaccine (TIV) against culture-confirmed influenza in healthy adults aged 18-64 years. A total of 4137 and 2066 subjects received TIV or placebo, respectively, during the 2005-2006 season in the Czech Republic. Our analysis included 632 and 315 subjects in the TIV and placebo groups, respectively.

2. Beran et al. (2009b) performed a randomized, double-blind, placebo-controlled study of the efficacy of TIV against culture-confirmed influenza in healthy adults aged 18 to 64 years. A total of 5103 and 2549 subjects received TIV or placebo, respectively, during the 2006-2007 influenza season in Czech Republic and Finland. Our analysis included 291 and 148 subjects in the TIV and placebo groups, respectively.

3. Jackson et al. (2010) performed a randomized, double-blind, placebo-controlled efficacy study of TIV against culture-confirmed influenza in healthy adults aged 18-49 years. In this study, a total of 3783 and 3828 subjects received TIV or placebo, respectively, during the 2005-2006 and 2006-2007 influenza seasons in the US. Our analysis included 1298 and 216 subjects in the TIV and placebo groups, respectively.

4. The Influence 65 trial (McElhaney et al. 2013) was a randomized, observer-blinded study of the relative efficacy of an adjuvanted vaccine (AS03-TIV) versus TIV against PCR-confirmed influenza in healthy adults aged 65 and over. The study included 43695 subjects from 15 countries who received AS03-TIV or TIV during the 2008-2009 and 2009-2010 seasons. The immunogenicity subset included 2422 and 2408 subjects in the AS03-TIV and TIV groups, respectively, and this analysis included immunogenicity data from the 2008-2009 season.

Despite some specificities, the four trials followed the same general protocol. During the study periods, subjects were monitored for influenza-like illness (ILI) by active surveillance (telephone contact/study center visit/home visits by study personnel), and by passive surveillance whereby subjects notified the study center if they experienced

ILI symptoms. Nasal and throat swabs were obtained from subjects reporting ILI. Laboratory identification of influenza viruses and case definitions in each study have been previously described (Beran et al. 2009b,a; Jackson et al. 2010; McElhaney et al. 2013). In all studies, blood samples were taken before vaccination and 21 days after vaccination to assess the level of serum antibodies in the immunogenicity sub-cohorts subjects.

3.1.2 Statistical analysis

Descriptive analysis and modelling were performed on the four randomized trials. In the descriptive analysis, the distribution of the following variables was characterized: gender, age, seasonal influenza vaccination history within previous 2 years, A/H3N2 infection status by the end of the study season, pre- (Day 0) and post-vaccination (Day 21) HI antibody titres against A/H3N2, pre-vaccination A/H3N2 seroprotection status (HI titre $\geq$1:40), vaccine received (AS03-TIV or TIV), and season strength ("strong season" or "low/moderate season"). In Beran et al. (2009b,a) and in Jackson et al. (2010), the season strength was based on the WHO influenza surveillance FluNet (Flahault et al. 1998) database and by evaluating the magnitude of the epidemics in the corresponding countries at the time the studies were conducted. In the Influence 65 trial (McElhaney et al. 2013), season strength was based on national surveillance data and attack rates in the study, as assessed by the Adjudication Steering Committee for the influenza peak season , which included influenza research experts independent of study sponsor. For continuous variables, the number of observations, mean, standard deviation, and minimum and maximum values were computed. For HI antibody titres, GMTs and their coefficient of variation were also calculated after a log10 transformation. Frequency statistics, including counts and proportions were obtained for the categorical variables. A preliminary graphical analysis was performed to assess each covariate versus A/H3N2 attack rates. The proportion of subjects with laboratory-confirmed A/H3N2 influenza was calculated for each dilution factor of the post-vaccination HI antibody response against A/H3N2.

In the modelling part of the analysis, a logistic regression model was used to assess the effect of exploratory variables on A/H3N2 disease occurrence. The following covariates were considered: pre-vaccination immunity state (titre greater or equal to 1:40 defined as "protected"), Day 21 post-vaccination A/H3N2 log 10 titres, gender, history of vaccination (vaccination 1 and 2 years before study start), age (Influence 65 trial $\geq$65 years versus other trials <65 years), and season strength. A manual stepwise variable selection was performed based on the Bayesian information criterion (BIC) to select the best comb. ination of covariates to describe the disease occurrence. The Akaike Information Criterion and Bayesian Information Criterion are both penalized-likelihood criteria for model selection (Burnham and Anderson 2004). The BIC tends

to select more parsimonious models compared to the AIC, which renders it preferable in the specific context of COP.

A receiver operating characteristic (ROC) curve was derived, presenting the sensitivity against one minus the specificity at various A/H3N2 HI antibody titre cut-off values (see note 3.1.2). The Youden index (Youden 1950) was used to identify the lowest titre at which the sum of the specificity and sensitivity was maximum. The Youden index can be interpreted as a cut-off titre between protection and no protection: sensitivity was defined as the proportion of subjects with confirmed A/H3N2 influenza and post-vaccination titres below the cut-off value; specificity was defined as the proportion of subjects in whom A/H3N2 was not confirmed and whom had post-vaccination titres equal to or greater than the cut-off value. Because the non-cases included both protected subjects and not sufficiently exposed/non-protected subjects, we also derived a cut-off post-vaccination titre value giving more weight to the cases detected, which we defined as the HI antibody titre cut-off values for the detection of A/H3N2 influenza with 90% sensitivity.

Note on the building of a ROC curve

The table below summarizes the possible outcomes of a test for detecting an event of interest by crossing the true disease status (infected or not, column-wise) with the test outcome (below or over the threshold of protection defined based on the logistic regression model, row-wise). True positive occurs when the post-vaccination HI titre is below the threshold of protection for a subject infected during the influenza season, true negative when the post vaccination titre is over the threshold for a non-infected subject.

Diagnosis based on the threshold of protection	**Reality**	
	Not infected (I-)	**Infected (I+)**
Not protected ($<$ threshold)	False positive (FP)	True positive (TP)
Protected ($\geq$ threshold)	True negative (TN)	False negative (FN)

Two types of errors can occur. First, false positive happens when a subject with a post vaccination HI titre below the threshold, considered as not protected, did not get infected by the end of the season. Second, false negative occurs in the opposite situation: post vaccination titre over the threshold of protection, considered as protected, for a subject who got infected.

Specificity and selectivity are two popular indicators or the validity of a test for detecting correct diagnosis. They are the probabilities of making a correct diagnosis

among the infected subjects (I+) and the non-infected subjects (I-). In a ROC curve, the following quantities are presented for varying levels of threshold of protection.

$$\text{True Positive Fraction} = \text{Sensitivity} = \frac{TP}{I+} \tag{3.1.1}$$

and

$$\text{False Positive Fraction} = \text{1-Specificity} = 1 - \frac{TN}{I-} \tag{3.1.2}$$

3.1.3 Results

An overview of subjects included in the analysis is shown in figure 3.1. Since the elderly did not receive a placebo vaccine, as opposed to the young adults, there is less data available for this population at low post-vaccination HI titres. The demographic characteristics and geometric mean titres (GMTs) for subjects included in the analysis by trial are shown in table 3.2. Pre- and post-vaccination GMTs in subjects who received TIV or AS03-TIV were 10.5-17.4 and 131.7-285.6, respectively, and in subjects who received placebo were 13.1-15.4 and 13.3-15.8, respectively. At baseline, 5405 (71.1%) subjects had an antibody titre against A/H3N2 that was <1:40 and 2309 (29.9%) had a titre that was above 1:40. Sixteen subjects did not have pre-vaccination titre data available. The A/H3N2 infection rates by age and season strength are shown in table 3.3. The frequency of A/H3N2 cases and post-vaccination HI antibody titres against A/H3N2 are shown in figure 3.2. Among 1098/7730 (14.2%) subjects with post-vaccination HI titres below 1:40, 24/1098 (2.2%) subjects had confirmed A/H3N2 illness; among 6632/7730 (85.8%) subjects with post-vaccination titres above 1:40, 50/6632 (0.75%) had confirmed A/H3N2 illness.

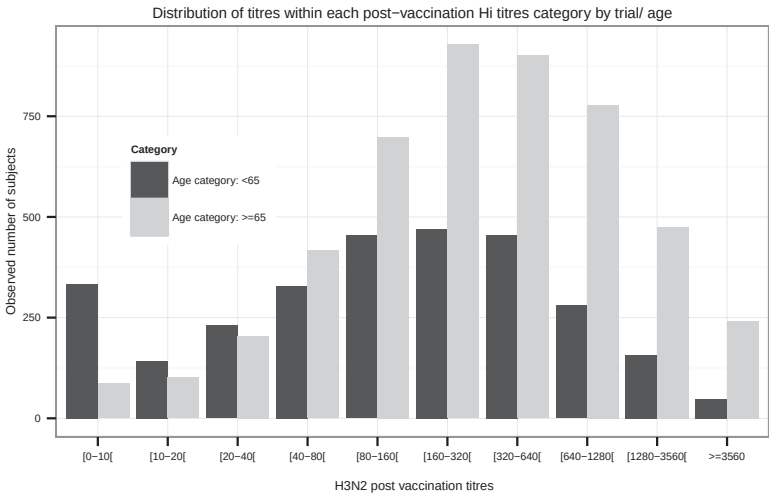

Figure 3.1: Distribution of the post-vaccination HI titres, by age/trial

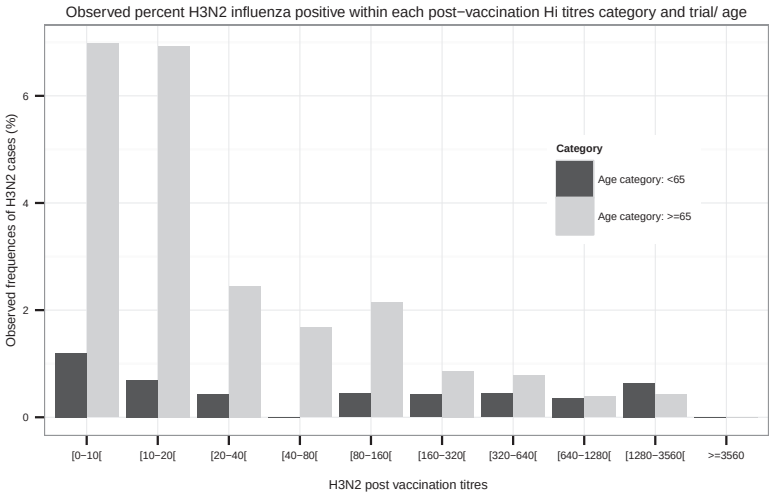

Figure 3.2: Observed proportions of influenza cases by post-vaccination HI titres category and age/trial.

	Trials in subjects <65		Influence 65 (≥65	
	TIV	**Placebo**	**AS03-TIV**	**TIV**
	N=2221	N=679	N=2422	N=2408
Mean (SD) range years	35.78 (12.2) 18-64	34.53 (11.3) 18-64	73.2 (6.0) 65-95	73.4 (6.3) 65-100
Vaccination history, n (%)				
1 year	243 (10.9%)	81 (11.9%)	1647/2197 (75.0%)	1650/2199 (75.0%)
2 years	187 (8.4%)	32 (4.7%)	1569/2119 (74.0%)	1578/2127 (74.2%)
HI GMT, (range)				
Day 0	14.03 (5-1810)	14.12 (5-640)	17.4 (5-1810)	17.4 (5-1280)
Day 21	178.61 (5-7240)	14.14 (5-905)	285.6 (5-20480)	172.3 (5-20480)

Table 3.2: Demographic characteristics and A/H3N2 HI antibody titres pre-vaccination (Day 0) and post-vaccination (Day 21)

Age / trial	Epidemic intensity	Subjects	A/H3N2 cases	Infection rate
$\geq$ **65**	**Low or moderate**	2939	20	0.68
	high	1891	40	2.12
$<$ **65**	**Low or moderate**	1873	4	0.21
	high	1027	10	0.97

Table 3.3: A/H3N2 infection rates by age and epidemic intensity in subjects pooled from four VE trials (immunogenicity sub-cohorts

	Parameter estimate	p-value	Odds ratio	95% CI on the odds ratio
Baseline risk	-3.92	<0.0001		
Post-vaccination log-titre	-1.20	<0.0001	0.33	0.23, 0.46
Epidemic intensity	1.24	<0.0001	3.44	2.10, 5.64
Age/trial	1.24	0.0001	3.45	1.89, 6.32

Table 3.4: Logistic regression parameter estimates and odds ratios for the selected model

The selected model includes three covariates: post vaccination log-titres, age category/trial and season strength. Table 3.4 shows the parameter estimates for the selected model. The odds ratio for A/H3N2 infection in high versus moderate/low season was 3.4 (95% CI: 2.1, 5.6). The odds ratio of A/H3N2 infection in subjects aged $\geq$ 65 years versus <65 years was 3.03 (95% CI: 1.9, 6.3). In our model, a four-fold increase in HI titre was associated with a 49.0% decrease in the risk of infection. Consistency of the HI response across the HI range was an assumption of the statistical model, which appeared acceptable based on the cases and HI titres observed (fig. 3.3).

The area under the curve for the ROC including age and season strength as covariates was estimated at 0.77 (Figure 3.4). It is significantly different from 0.5 ($p<0.0001$). The Youden index HI antibody titre cut-offs in subjects aged <65 years were 1:5 in a low/moderate season and 1:40 in a high season; the cut-offs in subjects aged $\geq$65 years were 1:40 in a low/moderate season and 1:640 in a high season. The 90% sensitivity HI antibody titre cut-offs in subjects aged <65 years were 1:28 in a low/moderate season and 1:453 in a high season; the cut-offs in subjects aged $\geq$65 years were 1:453 in a low/moderate season and 1:5120 in a high season.

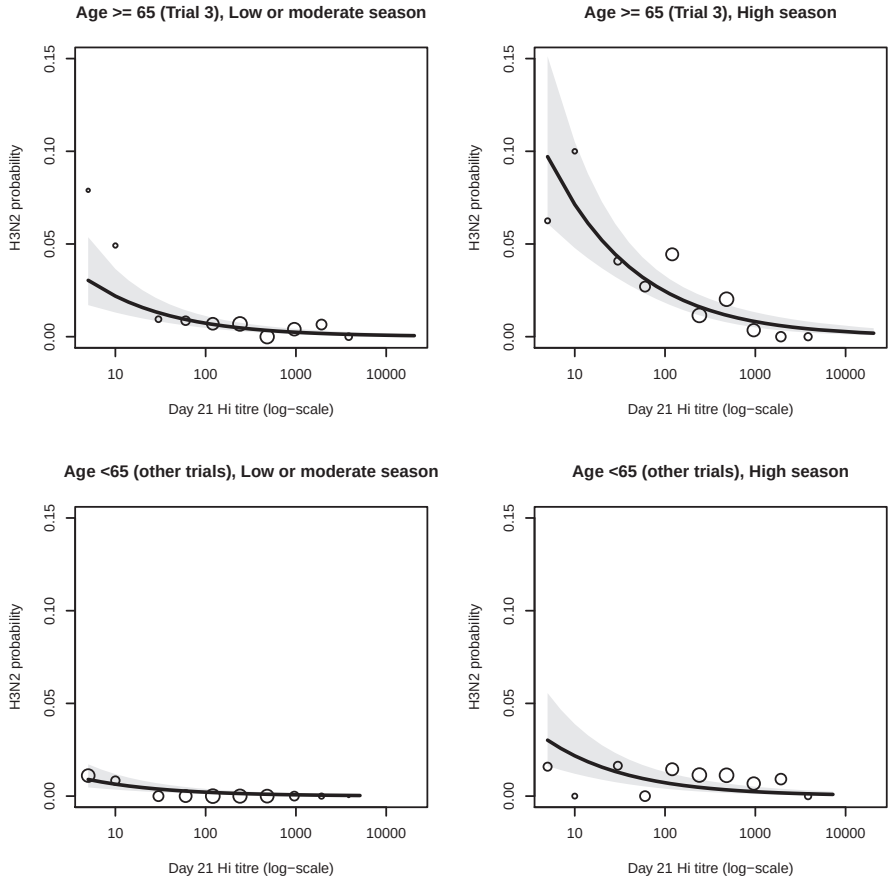

Figure 3.3: Estimated risk, point-wise 95% CI and observed proportions of clinical infections over post vaccination HI log10 titres, by age/trial and season strength. Note: the data point size represents the number of subjects for the post vaccination level of HI titre.

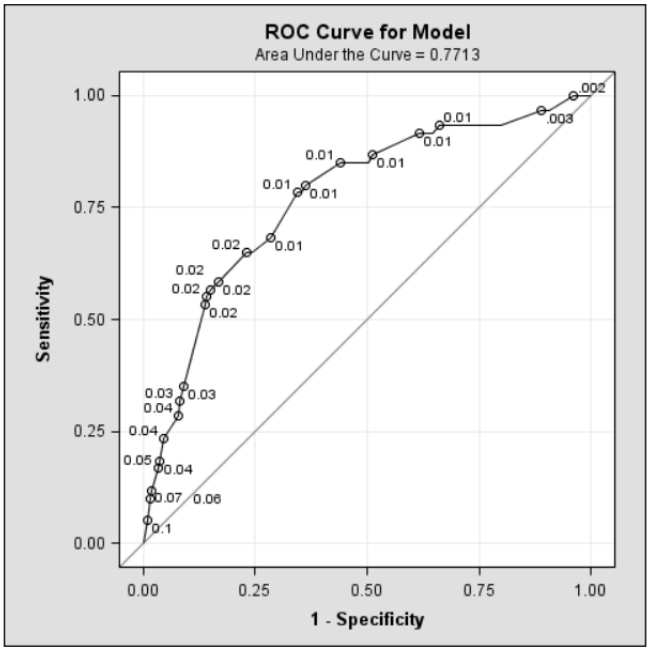

Figure 3.4: ROC curve for the selected model, including post-vaccination HI titres (log scale), subjects age category and season strength as covariates. The area under the curve is 0.7713 and is significantly different from 0.5 ($p<0.0001$)

3.2 Discussion

In our COP analysis of four randomized trials, two in adults aged 18-64 years, one in adults aged 18-49 years and one in adults aged $\geq$ 65 years, age influenced the baseline risk of A/H3N2 infection with an odds ratio of 3.4 for subjects aged $\geq$ 65 years versus <65 years. Including age and season strength as covariates, a four-fold increase in titre was associated with a two-fold decrease in the risk of A/H3N2 infection, with a similar difference in risk observed when including only age as a covariate. While age did not appear to affect the serological response to vaccination, older subjects appeared to have a greater risk of infection at similar titres compared with younger subjects. The Youden index cut-off values in a high season were 1:28 and 1:453 in subjects <65 years and $\geq$65 years, respectively.

Some factors that are confounded with differences between the trials: the nature of the comparison (placebo or active reference), influenza case definitions, laboratory methods for viral detection (culture or PCR), HI measurements, the age of participants (18-64 years, 18-49 years, $\geq$65 years), and the fact that the studies were conducted in different countries and seasons. In addition, cell-mediated immunity may impact protection against influenza, and the influence of mechanisms other than humoral immunity (based on HI titres), was not accounted for in our analysis. We believe that the most important differences among the trials were age, the endpoint that considered all cases of A/H3N2 infection as relevant despite the emergence of drift strains, and the fact that in the $\geq$65 years study, subjects were not classified as vaccine-matched.

Influenza occurrence first depends upon the exposure of a given population to circulating viruses. Because countries have diverse vaccination policies that may influence the transmission and exposure to influenza, we included season strength as a covariate based on surveillance in each country as an indicator of exposure. However, exposure may change the level of antibody needed to prevent illness of any severity; this is an important concept, because in adults, most illnesses are relatively mild, but the risk of severe illness resulting in hospitalization and adverse outcomes increases with age. In this pooled analysis, we did not have systematic prospective classification of moderate to severe illness, and as the antibody level correlating with protection against moderate to severe illness may be lower than that required to protect against mild illness, this study may be confounded. Indeed, differences in vaccine efficacy arising from regional and seasonal variability of multiple circulating viruses (with wide heterogeneity in prevalence and variations in the severity of influenza illness) are difficult to account for in a correlates of protection analyses. This point will be discussed in more details in the context of phase III VE trials in Chapter 6.

An objective of our analysis was to try to identify the HI titre that best separates the two subject groups - protected and not-protected. By considering a ROC approach, selecting a cut-off point involves a trade-off between sensitivity (the probability of

a case being not-protected by having a titre below the threshold) and specificity (the probability of a non-case being classified as protected by having a titre value above the threshold). The Youden index gives the same weight to both sensitivity and specificity as it defines the cut-off point as the titre value that maximizes the sum of the sensitivity and specificity (Youden 1950; Kelly et al. 2008).

The Youden index method depends upon the separability of the protected and non-protected populations. However, the rate of infection among subjects with low titres may be strongly associated with the chance of exposure and disease prevalence, which vary between seasons and locations and social behaviour. Therefore, the HI titre density curves for two subpopulations, the first comprised of subjects who were not infected are the results of a mixture of subjects who were protected and the second comprised of subjects who were not protected but were not sufficiently exposed to be infected. Several methods have been proposed to account for this issue. First the methodology we used relies on the belief that false negatives are likely to occur and thus sensitivity (true cases) should determine the cut-off value. As well as Youden index in our study we reported the cut-off for 90% sensitivity, which was 453 in subjects <65 years and 5120 in subjects aged $\geq$65 years in a high season.

Another analytical method, suggested by Dunning (2006), uses scale logistic regression modelling in which the probability of the subject developing influenza is the probability that the subject is susceptible multiplied by the probability that susceptible individuals develop disease (Coudeville et al. 2010). In addition, Li et al. (2013) developed a dichotomization method based on the maximisation of the correlation between the two populations and the dichotomous variable. In the non-cases population, the methods included a parameter defining the probability that the observation arises from the case population (unprotected but not exposed).

In our logistic regression model, we found that the baseline risk of disease was higher for older than younger subjects and was higher in strong season than a moderate/low season. Subjects with a greater baseline risk (i.e. older subject and/or in a strong season) will need higher antibody titres to have the same level of protection as subjects with a lower baseline risk, meaning that our model provides varying cut-off points for protection. In our model, we did not find a significant interaction between post-vaccination titres and subject-related covariates, although given the lack of power, the results should be interpreted with caution. However, we did find as reported in the literature that a four-fold increase in post-vaccination titres was associated with a two-fold decrease in the risk of infection.

3.3 Conclusion

While post-vaccination HI titres cannot be interpreted as a surrogate of protection as defined by Qin et al. (2007), there seems to exist a positive consistent correlation between this measure and the risk of symptomatic influenza. Thus, within a homogeneous population, people with higher post vaccination HI titres appear to be better protected than those with lower HI titres. However, this relationship does not seem to hold the setting of an heterogeneous population. In our analysis, we found that age and the levels of exposure were related to varying risks of disease depending on the post vaccination HI titres. Although HI titres seem to be a useful COP to help develop a vaccine, they do not have a good predictive power for the occurrence of the disease and should not be interpreted as useful in this manner. Finally, throughout the clinical development of a new vaccine, we recommend the use of relative immunogenicity protection criteria, i.e. seroconversion factors and rates instead of seroprotection rates.

In the next chapters, we will focus our attention to the third phase of the clinical development of new vaccines.

Summary of Chapter 3

- Vaccine development against seasonal influenza is based on post vaccination HI titres. However, an experimental vaccine formulated to increase the immunogenicity response has failed to show improved efficacy compared to a standard of care
- Through the assessment of HI response as a COP in a post-hoc analysis of a pooled dataset of four VE trials, we confirm the correlation between post vaccination HI titres and risk of developing clinical infection
- We also found that the baseline risk of clinical disease depended upon the age/trial and season strength taken as an indicator of exposure to the virus. Consequently, we believe that a universal threshold of protection does not seem achievable since the protection level depends upon other factors, such as subjects characteristics and level of exposure. Relative immunogenicity protection criteria, i.e. seroconversion factors and rates instead of seroprotection rates, should therefore be preferred in the immunogenicity trials.

Chapter 4

Influenza vaccine efficacy trials: a simulation approach to understand failures from the past

This chapter is mainly based on the paper "Influenza vaccine efficacy trials: a simulation approach to understand failures from the past" by Benoit A., Legrand C. and Dewé W., conditionally accepted by Pharmaceutical Statistics.

The success of a seasonal influenza vaccine efficacy trial depends upon the design but also the annual epidemic characteristics. In this context, simulation methods are an essential tool in evaluating the performances of study designs under various circumstances.
In this chapter, we discuss why traditional methods for simulating time-to-event data are not suitable for the simulation of influenza vaccine efficacy trials. Instead, we propose a mathematical model parameterized with historical surveillance data, heterogeneous frailty among the subjects, survey-based heterogeneous number of daily contact and a mixed vaccine protection mechanism.
We illustrate our methodology by generating multiple trials data similar to a large phase III trial that failed to show additional relative vaccine efficacy of an experimental adjuvanted vaccine compared to the reference vaccine. We show that small departures from the designing assumptions, such as a smaller range of strains protection for the experimental vaccine or the chosen endpoint, could lead to smaller probabilities of success in showing significant relative vaccine efficacy.

4.1 Introduction

In Chapter 3, we showed that while post vaccination HI titers do not predict the efficacy of a vaccine there was still a correlation between the immunogenicity response and protection against clinical influenza. Adjuvanted vaccine thus, formulated to increase the immunogenicity response, should provide increased VE compared with a standard vaccine. Unfortunately this could not be confirmed in the large Influence 65 trial. In the season 2008-2009 this large, multi-countries trial in the elderly ($\geq$65 years) failed to show improved efficacy against all influenza A and B strains of a new adjuvanted vaccine versus the standard vaccine (McElhaney et al. 2013). Because of the study population, no placebo group could be included in the trial and all subjects were vaccinated, either with the experimental vaccine or with the standard one.

Possible reasons for the failure of this trial include:

- The endpoint chosen for this trial was too ambitious: it had been assumed that the adjuvanted vaccine would give protection against the vaccine strains but also against other strains (excluding the pandemic strain)
- The absolute efficacy of the new vaccine was lower than expected while the absolute efficacy of the reference vaccine was as expected, resulting in a lower relative efficacy defined as the efficacy of the new vaccine compared to the reference vaccine

In the recent years, multiple failures of influenza VE trials have been observed (Ohmit et al. 2008; Beran et al. 2009b; Jackson et al. 2010; McElhaney et al. 2013; Tsang et al. 2014). To learn from past failures and improve influenza vaccine development, it is essential to understand what might have happened. Indeed, when designing and analysing VE trials, complex issues related to particularities of influenza have to be dealt with (see Chapter 2). Influenza viruses are constantly evolving, which makes the intensity of the the seasonal epidemics and their predominant circulating strains somewhat unpredictable (Gupta et al. 2006). As a lot of complex factors must be properly accounted for, sound, detailed, evidence-based planning is required to design and conduct powerful clinical trials able to demonstrate efficacy of an experimental vaccine. Therefore clinical trials may fail to show significant VE for reasons other than lack of efficacy, for example because of low virus circulation or mismatch between the circulating and vaccine strains. In such a complex context, simulation studies are a particularly useful tool to understand failure of past trials, to investigate the performance of various designs and ultimately to help design high-quality trials (Burman et al. 2005; Burman and Wiklund 2011).

Since the most fundamental piece of information collected in a VE study is the time of onset of each influenza episode, simulating VE data means generating time-to-event

data. Classical data generation techniques for time-to-event data are usually based on simple models, such as the Cox proportional hazards (PH) model, assuming a parametric baseline function (Bender et al. 2005). However, we will show that such models cannot appropriately take into account the fact that influenza yearly epidemics depend on many factors. The particularities of influenza, such as seasonality, heterogeneity of individuals and their level of exposure (Longini Jr and Halloran 1996), virus circulation, mechanisms of vaccine protection (Smith et al. 1984; Halloran et al. 2010) and regular mismatches between vaccine and circulating strains are not captured by these models. We therefore propose a mathematical model inspired from the epidemiological literature (Chao et al. 2010). Our objective is to generate data consistent with the predominant characteristics of seasonal influenza to assess trial designs and data analysis methodologies.

Section 4.2.1 describes characteristics of influenza are not reflected in classical data generation methods. In section 4.2.2, we propose a new model and a new algorithm allowing to take them into account.

Based on this, a simulation study is performed in Section 4.3 to show how it can help better understanding the failure of the trial mentioned above. We simulate data for similar trials, but adding a placebo group, and use these to investigate the impact of the choice of the trial main endpoint, the range of protection of the new vaccine, the presence of an immune portion in the population and of the real efficacy levels, absolute and relative.

4.2 Background and context

4.2.1 Simulation state-of-the-art and particularities of influenza

In this section, we describe the particularities of influenza and its spreading dynamic. We explain why these are not taken into account in the classical simulation time-to-event models and how we include those characteristics in our proposed model. We have selected influenza characteristics that appear to be relevant in regards to the outcome of VE trials. Our objective is to build a data generation algorithm that can be used to design a future VE trial. We intend our model to be relatively sraightforward to use while remaining flexible enough to reflect various scenarios.

Several epidemiological models have been developed for infectious diseases. Their main objective is to understand the spread of the diseases and the measures, such as vaccination, needed to control their propagation. Among those models, we cite the Susceptible-Infectious-Recovered (SIR) model (Coburn et al. 2009), a compartmental model, and FluTE (Chao et al. 2010), a stochastic model specific to influenza epi-

demics. None of those models however can be directly applied to generate clinical trials data such as needed in our work. Indeed one of the main specificity of simulating clinical trials data is that the trial population is homogeneous but has contacts with an heterogeneous population. As a result, a large population would have to be generated through an epidemiological model and the trial participants would then be sampled within the sub-population of interest whereas our methodology allows the direct simulation of only the trial participants.

Cox PH model

In the context of influenza VE trials, the time-to-event represents the time between vaccination and event or censoring. Simulating influenza VE trials requires a model: the parametric proportional hazard regression (Cox 1972) is frequently used to simulate data. Any covariate is supposed to have a multiplicative effect with respect to the baseline hazard function and, given the covariates, subjects are assumed to have the same risk of event.

Seasonality and geographical regions heterogeneity

Influenza infections are characterized by seasonality. The clinical influenza incidence curve over time is almost flat during most of the year and present one or more peaks between November and March in the Northern Hemisphere and between May and October in the Southern Hemisphere (Figure 4.1)(Lofgren et al. 2007).

Traditional parametric survival time distributions, such as exponential, Weibull and lognormal densities are characterized by one or two parameters (Burton et al. 2006). Those distributions are characterized by a constant or unimodal hazard function. However, identifying the distribution that would fit the timing and intensity of the peak is neither flexible nor straightforward. Three-parameters extension of those distributions (Cooray 2006; Reed 2011) and polyhazards models (Tsai et al. 2013) have been developed to improve the flexibility of the hazard functions, allowing multimodal hazards. The piecewise exponential distribution (Demarqui et al. 2008; Kim and Proschan 1991) is another flexible option for modelling time-to-event data. The difficulty in generating data when using these flexible parametric distributions lies in the selection of the parameters and in the case of the piece-wise model, the determination of the hazard change points.

Therefore, in this context, a parametric baseline hazard shape is not recommended to generate data. Instead, we propose to use incidence historical data, available in epidemic surveillance databases such as FluNet (Flahault et al. 1998). This methodology will be detailed in Section 4.3.

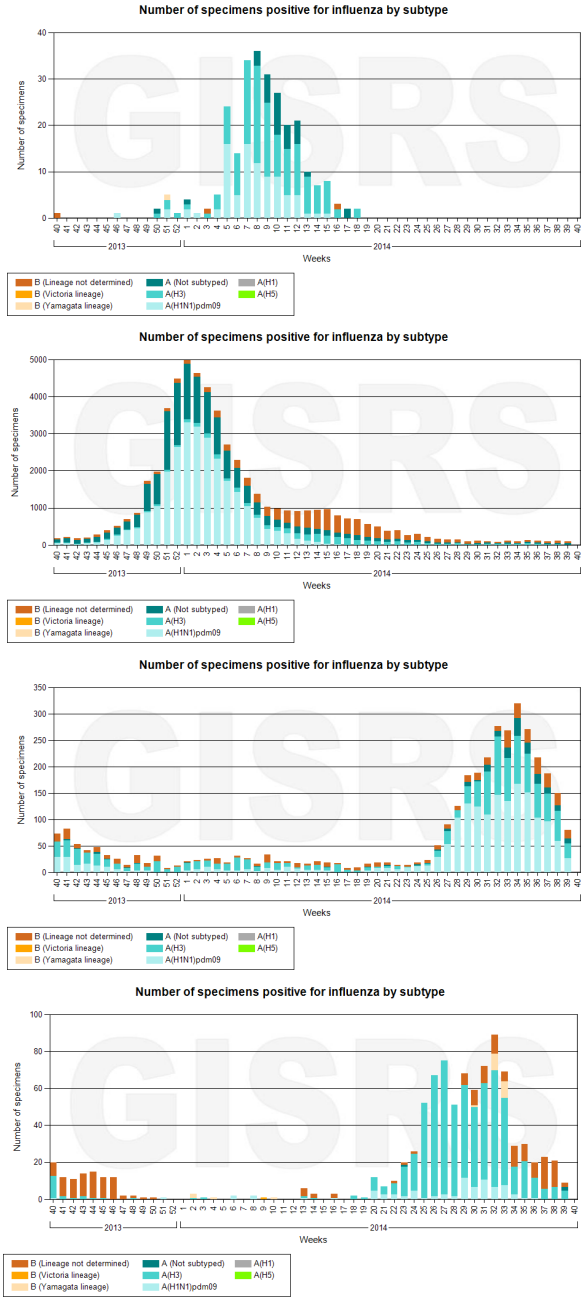

Figure 4.1: 2013-2014 flu seasons in four countries: at the top, two from the Northern hemisphere (Belgium and US) and at the bottom, two from the Southern hemisphere (Australia and South Africa). The colors of the bars represent the virus strains. From the FluNet website.

This proposal makes the inclusion of the multiplicity of the strains causing yearly epidemics straightforward. Indeed since influenza is not a single, genetically stable virus: several strains co-circulate and the viruses mutate continuously. The repartition of the strains and their level of circulation differ between geographical regions, resulting in varying times of occurrence and magnitudes of epidemic peaks (Dewé et al. 2013). Generating data using a parametric baseline hazard function for different influenza sub-types and geographical regions would require the selection of parameters for each strain and for the assumed correlation structure between the regions. Instead, we use country specific and strain specific historical incidence data to take into account the geographical aspects of the circulation and the repartition of the strains.

Subjects fragility

Not everybody is equally affected by influenza (Chapter 2) while traditional methods of simulation for time-to-event assume an equal risk for all subjects given the covariates. In the seasonal influenza context, this is not verified as unobserved sources of heterogeneity, such as subject fragility, must be taken into account. We propose to include subject weakness and contact rate random effects through:

$$\lambda(t|Z = z, X_1 = x_1, X = x) = zh(t|X_1 = x_1, X = x) \tag{4.2.1}$$

where z is an unobservable realization of a non-negative random variable with a given probability density function f_Z. The random variable z is often called a frailty (as people with a higher value will have a higher hazard, and thus be more "frail") and model (4.2.1) is then referred to as the frailty model (Duchateau and Janssen 2007).

Contacts

Influenza infection is conditional to virus exposure, i.e. contact with an infected person. The higher the number of contacts, the higher the risk of infection. The daily contact rate can be modelled from a discrete distribution. Selecting a distribution admitting mass at 0 allows for the inclusion of subjects who are not exposed at all. For the purpose of our model, we consider that the number of daily contacts for one individual is constant across time.

Vaccine protection mechanisms

The PH model actually assumes that all the vaccinated individuals are equally protected against the virus via the factor $\exp(\beta_1)$ (Equation 2.2.10 in chapter 2) acting

multiplicatively on the baseline hazard. However, the existence of two mechanisms of vaccine protection, *leaky* and *all-or-none*, has been argued (see Chapter 2 for more details).

Arguing that vaccine protection is likely to be a mixture between both *leaky* and *all-or-none* mechanisms, we introduce the two protection mechanisms in our model (Figure 4.2). Consequently, individuals from the experimental group have either a complete or a partial protection against the studied viruses. Furthermore, we consider that some subjects could be previously immune to the infection, for example because of prior vaccination or influenza infection with a genetically closely related virus strain. We consider that a portion π_0 ($\pi_0 \geq 0$) of the subjects from the reference group is completely protected. This fraction is increased by an additive efficacy term VE_π in the experimental group resulting in a fraction π_1 of non-susceptible subjects in this group.

We therefore use two efficacy parameters (Halloran et al. 1996). First, VE_s represents the vaccine protection in the experimental versus the reference group in the susceptible population. Second, VE_π represents the proportion of the vaccinees totally protected that is induced by the experimental vaccine. The proportion of subjects totally protected in the experimental group is computed as $\pi_1 = \pi_0 + VE_\pi$ where π_0 is proportion of subjects totally protected in the reference group. Total efficacy is computed as

$$1 - \frac{(1 - VE_s)(1 - \pi_1)}{1 - \pi_0} \tag{4.2.2}$$

In the case of a *leaky* vaccine, $\pi_1 = 0$ and VE resumes to VE_s. If the vaccine only act as an *all-or-none* protection, $\text{VE}_s = 0$ and VE simplifies to $1 - \frac{1-\pi_1}{1-\pi_0}$.

4.2.2 Simulation Framework

Based on the previous points, we generate times of disease onset through a mixture cure model based on the principle that there are three conditions for any subject to get influenza: first he/she must be susceptible, second he/she must enter in contact with an infected person, and third transmission of the virus must occur (Halloran et al. 1991).

We consider a sample of n independent individuals participating in a phase III VE trial. The subjects are potentially exposed to the infectious agent of interest over a period of time $[0, T]$, where T is the end of the surveillance period. So T is also a censoring time for the subjects without influenza episode. The model could easily be extended to the case of non-informative random censoring. Subjects are randomized to receive either the experimental or the reference vaccine.

In our model, a subject i ($i = 1, ...n$) has contact with other persons at a rate of c_i contacts per unit of time. The probability at time t that a contact is infectious for

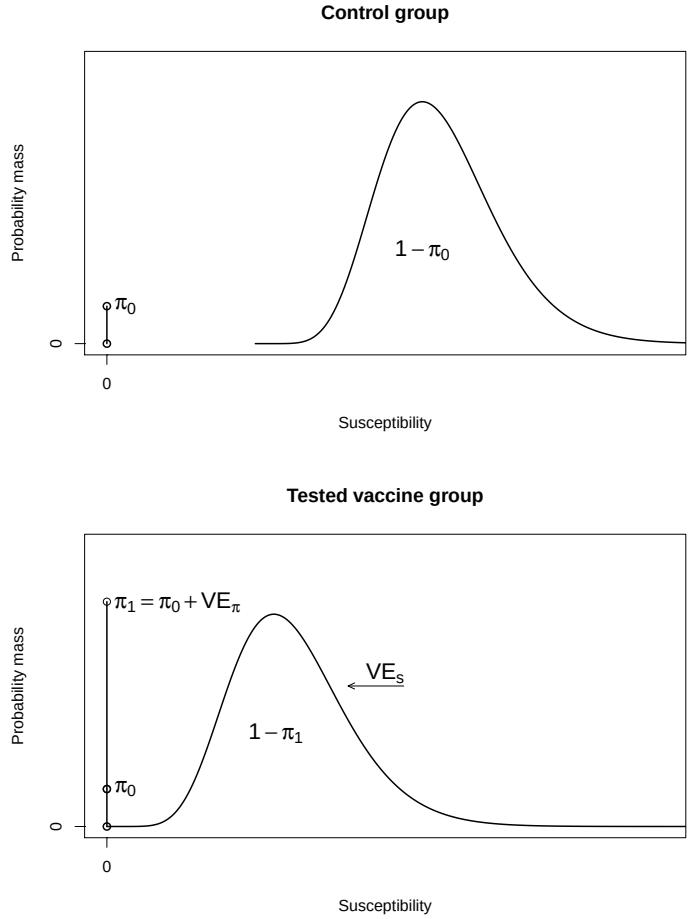

Figure 4.2: Illustration of Mechanisms of Vaccine Efficacy (adapted from Halloran et al. (Halloran et al. 2010)): distribution of susceptibility in the control (top) and new vaccine (bottom) groups of subjects. A fraction π_0 of the subjects are completely protected in the reference (placebo or active control) group. This fraction is increased by VE_π in the experimental group, resulting in a fraction of totally protected subjects of π_1. The level of susceptibility of the remaining $1 - \pi_1$ experimental group subjects is reduced by a factor VE_s compared to the reference group ones.

influenza strain k is defined by the prevalence of the disease caused by strain k at time t, $p_k(t)$, $k = 1, ...K$. If a susceptible reference group person makes a single contact with a person infected with the pathogen of interest, then he/she becomes infected with probability ρ, which is the transmission probability to a reference group person. If a person from the experimental makes a single contact with an infected person, then that individual becomes infected with rate $(1 - VE_s)\,\rho$. The fragility term for subject i is noted z_i. It acts as a multiplier of the instantaneous hazard. To simplify the notation we consider a $1:1$ ratio and, for each vaccine group, i goes from 1 to $\frac{n}{2}$. So, the instantaneous hazards $h_{0,k,i}(t)$ and $h_{1,k,i}(t)$ for strain k at time t for the reference and experimental group subjects respectively are:

$$\begin{cases} h_{0,k,i}\,(t) = (1 - \omega_{0,i}) z_i c_i \rho\, p_k\,(t) \\ h_{1,k,i}\,(t) = (1 - \omega_{1,i})\,\{(1 - VE_s)\, z_i c_i \rho\, p_k\,(t)\} \end{cases} \tag{4.2.3}$$

$i = 1, ...,\ \frac{n}{2}$ with $g = 0,\ 1$ for the reference and the experimental group respectively, $t = 1, ...,\ T$, the protection status $\omega_{0,i} \sim$ Bernouilli(π_0) for the reference group and $\omega_{1,i} \sim$ Bernouilli$(\pi_0 + VE_\pi)$ for the experimental group.
The overall survival function for subject i, $S_{g,i}(t)$, for any strain, can be derived as:

$$\begin{cases} S_{0,i}\,(t) = \exp\left(- (1 - \omega_{0,i}) \left(z_i c_i \rho \sum_k \left(\int_0^t p_k\,(u)\, du \right) \right) \right) \\ S_{1,i}\,(t) = \exp\left(- (1 - \omega_{1,i}) \left((1 - VE_s)\, z_i c_i \rho \sum_k \left(\int_0^t p_k\,(u)\, du \right) \right) \right) \end{cases} \tag{4.2.4}$$

If we consider daily prevalences for the investigated strain(s), the integral of $p_k(u)$, i.e. the cumulative prevalence of the infection, is approximated through the cumulative sum of the daily prevalences of infection between time 0 and time t.

Unconditionally to the susceptibility status ω and for all circulating strains, we have for the comparator group

$$\begin{aligned}
&Pr(T \geq t|X_1 = 0) \\
&= Pr(T \geq t|\omega_0 = 0, X_1 = 0) \times Pr(\omega_0 = 0|, X_1 = 0) \\
&\quad + Pr(T \geq t|\omega_0 = 1, X_1 = 0) \times Pr(\omega_0 = 1|, X_1 = 0) \\
&= Pr(T \geq t|\omega_0 = 0, X_1 = 0) \times (1 - \pi_0) + \pi_0 \\
&= \exp\left[-\int_0^t (h(u|\omega_0 = 1, X_1 = 0)du)\right] \times (1 - \pi_0) + \pi_0 \\
&= \exp\left[-z_i c_i \rho \int_0^t (p(u))du)\right] \times (1 - \pi_0) + \pi_0
\end{aligned} \tag{4.2.5}$$

and for the experimental group

$$\begin{aligned}
&Pr(T \geq t|X_1 = 1) \\
&= Pr(T \geq t|\omega_1 = 0, X_1 = 1) \times Pr(\omega_1 = 0|, X_1 = 1) \\
&\quad + Pr(T \geq t|\omega_1 = 1, X_1 = 1) \times Pr(\omega_1 = 1|, X_1 = 1) \\
&= Pr(T \geq t|\omega_1 = 0, X_1 = 1) \times (1 - \pi_1) + \pi_1 \\
&= \exp\left[-\int_0^t (h(u|\omega_1 = 1, X_1 = 1)du)\right] \times (1 - \pi_1) + \pi_1 \\
&= \exp\left[-(1 - VE_s) z_i c_i \rho \int_0^t (p(u))du)\right] \times (1 - \pi_1) + \pi_1
\end{aligned} \tag{4.2.6}$$

We can see that for $t \to +\infty$ the marginal survival functions converge respectively to π_0 and π_1 in the comparator and the experimental vaccine groups.

4.3 Illustration

Our simulation methodology is very flexible and can be used in many situations. For example, through the strain-specific capacity of our simulation framework, the benefit of adding a fourth strain in the seasonal vaccines could be assessed. It is also a powerful tool for designing trials since it allows to generate multiple scenarios and to test the robustness of the chosen design and statistical analysis model in different situations.

In the next section, we illustrate how our simulation model is used as a retrospective tool. We show how it can be used to better understand the outcome of the trial mentioned earlier (and described in more details below). Simulations will be used to investigate the likelihood of three possible scenarios (primary endpoint too ambitious,

higher VE for the control than expected, lower VE for the experimental vaccine than expected) and their potential impact on the results. To do so, we will simulate data for trials designed like the original trial and ran during the same season and in the same countries but we will artificially include patients in a placebo group in these trials.

4.3.1 Original trial

The trial of interest included 43802 subjects aged over 65 from 15 countries worldwide during the 2008-09 influenza season. Subjects were randomly assigned (1:1) to receive either an adjuvanted new trivalent vaccine or the standard non-adjuvanted trivalent vaccine. The observation period started on November 15, 2008 and ended on April 30, 2009. The primary endpoint was the relative efficacy of the new vaccine versus the standard vaccine for the prevention of laboratory confirmed influenza A and B. The study protocol foresees a Cox regression analysis model including the covariates vaccine and country.

The trial was powered based on a power of 90% to confirm the relative vaccine efficacy with a one-sided type 1 error of 2.5% for an expected cumulative incidence of clinical infection during the observation period, i.e. AR of 1% in the group receiving the standard vaccine. It was assumed that the absolute efficacy for the new vaccine was 65% and 50% for the standard vaccine (relative efficacy of 30% for the new versus the standard vaccine).

This trial is registered with ClinicalTrials.gov, number NCT00753272 and has been published recently (McElhaney et al. 2013).

4.3.2 Simulated trials

We implemented our simulation methodology in SAS v.9.3. Data analysis can be done in SAS or in R v3.0.1 within the IML SAS to R capacity.

Our data generation algorithm is summarized in figure 4.3. All model parameters are combined according to equation 4.2.4, resulting in n individual survival curves. A random $[0, 1]$ uniform number u_i is generated for each subjects and disease occurrence time is defined as $min\{t : \ u_i \leq S_i(t)\}$. Observations are censored at the end of the trial. We consider that once a subject has been infected by any strain, he/she is immune for the rest of the season.

Contrarily to the traditional methodology, our simulation method requires the generation of individual survival curves. However, if one simulates the data effectively, the process is neither time nor memory consuming.

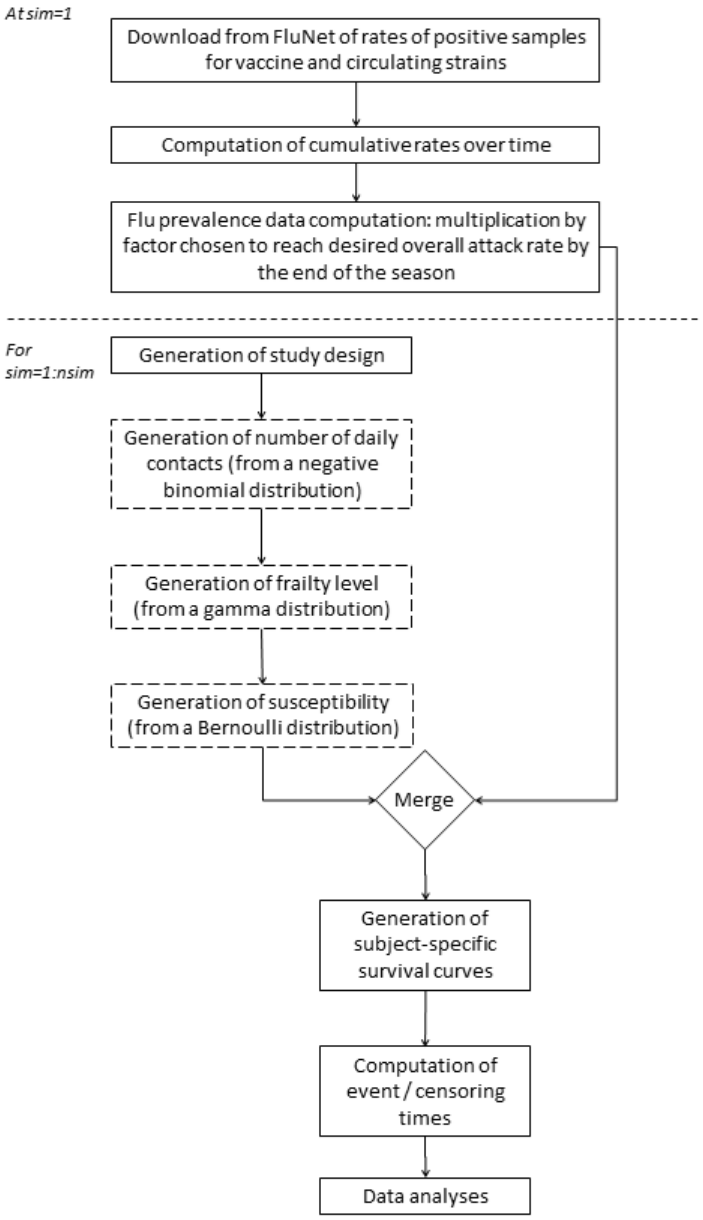

Figure 4.3: Data generation algorithm. Steps above the dashed line are performed once while steps below are performed for each simulation.

Trial design

We simulate trials over a single season including 43802 subjects (as in the original trial) plus 21901 subjects for the new placebo group, i.e. 65703 subjects equally allocated between 3 vaccine groups (placebo, standard vaccine and adjuvanted vaccine). Country representation and ages of the subjects are similar to the characteristics of the original trial.

Absolute VEs are defined as the hazard reductions between each vaccinated groups and the placebo group, estimated through a Cox regression model with vaccine and country as covariates. Relative VE is the risk reduction between the adjuvanted vaccine group and the standard vaccine group.

For each scenario, 500 trials are generated.

Note: number of simulations

Following Burton et al. (2006), we calculated the number of simulations required to obtain estimates within 5% accuracy of the assumed true value for the vaccine effect, β_1. The number of simulations to perform, B, is calculated as

$$B = \left(\frac{Z_{(1-\frac{\alpha}{2})}\sigma}{\eta} \right)^2 \tag{4.3.1}$$

where η is the specified level of accuracy from the true value β_1, $Z_{(1-\frac{\alpha}{2})}$ is the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution and σ^2 is the variance of the parameter of interest.

For an expected vaccine effect $\beta_1 = -0.69$ (VE=0.5), 100 expected events, $\sigma^2 = 0.04$ and $\alpha = 0.05$, the required minimum number of simulations is 125. With 250 and 500 simulations, the estimates obtained are within respectively 3.5% and 2.5% of accuracy of the assumed true value for the vaccine effect.

Vaccine efficacy

We consider seven combinations of efficacy levels of the two vaccines: the assumed scenario for the original trial design, two scenarios with lower than expected absolute efficacy for the standard vaccine and four scenarios in which the absolute efficacy of the standard vaccine is as expected but the efficacy of the new vaccine is lower than assumed.

Vaccine protection tends to be smaller for elderly compared to young adults and in this age group total protection is very unlikely (McElhaney et al. 2006). For this reason,

Scenario	Standard vaccine	Adjuvanted vaccine	Cases considered for the computation of VE
Mixed	Vaccine strains	All strains	All cases
Trivalent all cases	Vaccine strains	Vaccine strains	All cases
Trivalent matching cases	Vaccine strains	Vaccine strains	Matching vaccine strains

Table 4.1: VE scenarios simulated in terms of vaccine protection and cases considered

we only consider a leaky protection mechanism. In a first setting, we consider that all subjects in the control group are susceptible ($\pi_0 = 0$) and in a second setting we consider that because of their age thus prior influenza experience, the population of interest includes 20% of immune subjects.

The two vaccines administered in the original trial contain three influenza strains: a H1N1 A strain, a H3N2 A strain and a Yamagata lineage B strain. We consider that the standard vaccine only gives protection against those three strains. For the adjuvanted vaccine, we consider two cases: in the first one we simulate data considering that this new vaccine only protects against the strains it contains, as the standard vaccine. In the second one, we consider that the adjuvant additionally provided cross-protection against the strains not included in the vaccine (except the H1N1 pandemic strain that started circulating by the end of the season 2008-2009 in some countries). In the first case, we compute VE against all cases considered and VE against vaccines strains. Table 4.1 shows the three VE scenario simulated for all the levels of efficacy considered.

Virus circulation

We use historical data from FluNet (Flahault et al. 1998) for season 2008-09 for the original trial 15 countries.

We consider viral circulation to be proportional to the weekly percentage of tests positive for each strain. This quantity is divided by a constant calibrated to reach desired overall seasonal AR levels, here 2%. Daily strains prevalences are computed from the weekly incidences and average disease duration.

Daily contacts rate

Mossong et al. (Mossong et al. 2008) showed that the best model for contacts heterogeneity is a negative binomial density, characterized by mean μ and over-dispersion

parameter ϕ. They also found that numbers of contacts differed between countries and age categories. The probability of observing a non-negative number of daily contacts c_i for a subject i characterized by explanatory variables X_i follows a negative binomial distribution with mean μ varying with X_i and over-dispersion parameter ϕ. Over-dispersion parameter value is set as $\phi = \frac{1}{0.36}$ and values for the specific age category (≥ 65) and 15 countries mean numbers of daily contacts μ are derived from Mossong et al. (Mossong et al. 2008) whenever available. Number of daily contacts mean values for the other countries were made up in order to maintain variability in contacts rates in our samples.

Frailty

Fixed effects estimations have been shown to be robust to frailty strictly-positive distributions mispecification in a frailty model (Pickles and Crouchley 1995). The choice of the frailty distribution is therefore not of particular interest here. We make the classical choice of the continuous strictly positive one-parameter gamma distribution:

$$f_Z(z) = \frac{z^{\frac{1}{\delta}-1} \exp(\frac{-z}{\delta})}{\delta^{\frac{1}{\delta}} \Gamma(\frac{1}{\delta})} \tag{4.3.2}$$

with Γ the gamma function and δ the variability parameter. We fixed $\delta = 4$ in order to reflect a 10-fold risk difference between the mean and the percentile 99 subjects, translating large fragility differences between healthy subjects and subjects with concomitant co-morbidities (Figure 4.4).

Transmission probability

In epidemiological models, the spread of a disease is usually characterized by the basic reproduction number R_0. It is defined as the number of secondary infections that arise from a typical primary case in a completely susceptible population (Truscott et al. 2011). R_0 is a combination of the number of contacts per unit of time per individual, here the mean number of daily contacts, c, the probability ρ of transmitting the infection per infectious contact and the mean duration τ of the infection period (Eq 4.3.3).

$$R_0 = c\rho\tau \tag{4.3.3}$$

We consider that a subject infected with influenza is contagious for a period of 7 days (CDC 2011) and that the average number of contacts for an infectious subject is 18

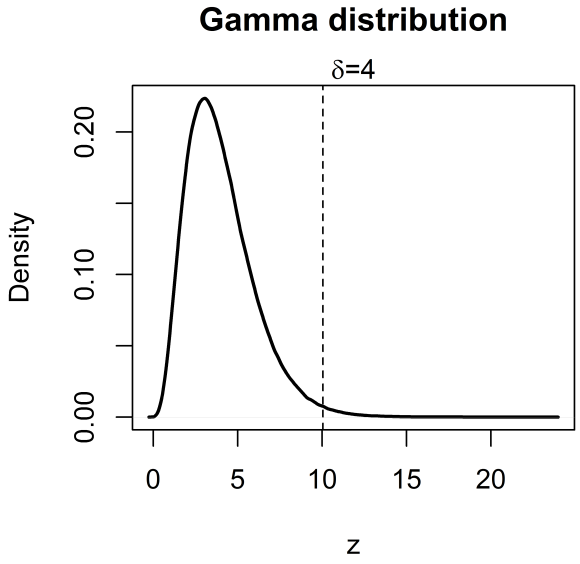

Figure 4.4: Gamma density with $\delta = 4$. The dashed line indicates the percentile 99.

(Mossong et al. 2008). The reproductive number for influenza has been estimated as ranging from 1 to 2, depending on the epidemic strength (Truscott et al. 2011). In order to reflect this information, a transmission probability $\rho = 0.01$ seems realistic.

4.3.3 Results

For each scenario, we estimate the VE of the adjuvanted vs. the standard vaccine through a Cox regression model with vaccine and country as covariates, as in the original trial. In table 4.2, we report the median estimate with 95% simulation confidence interval, defined as the $2.5 - 97.5$ inter-percentile range of the estimated values. We define the probability of success, P(success), as the proportion of simulated trials in which the lower limit of the 95% confidence interval is greater than 0.

In the season 2008-2009, influenza A H3N2 mainly circulated in the European countries, H3N2 and B strains in Canada, B strains in Mexico and the US and seasonal H1N1 in Asia (McElhaney et al. 2013). The Yamagata B strain was included in the vaccine, but mostly Victoria B strain (the other B lineage) circulated, which led to mismatch (Reed et al. 2012). Those characteristics are well captured in our simulations. The H1N1 pandemic strain also appeared at the end of the season in some

countries. However, as in the original protocol, it was decided not to take into account those cases in the analyses.

The original trial was designed assuming AR of 1% in the standard vaccine group. In our simulations, AR are calibrated for the placebo group. We obtain matching cases AR of 1.4% and 1.1% respectively when all subjects are susceptible and when 20% of them are immune. Similarly, when considering all cases (matching and non matching), AR of 2.2% and 1.7% are obtained in the placebo group without and with an immune fraction. Consequently, AR in the standard vaccine group under trial assumptions (absolute VE of 50% for the standard vaccine) are similar to the value of 1% used in the design step of the original trial.

In the situation where all trial assumptions are met (Mixed scenario in Table 4.2), computed relative VE results not only from higher efficacy of the new vaccine over the standard one for the vaccine strain (simulated relative VE, from 0 to 0.3) but also from the protection of this new vaccine against non vaccine strains (relative VE equal to the absolute VE). Median relative VE, simulation empirical 95% CI and probabilities of success (significant relative VE) are presented in Table 4.2. In this case, estimated relative VE against all strains is higher than the simulated relative VE against the three vaccine strains, resulting in higher probabilities of success to detect a significant gain of efficacy for the adjuvanted vaccine compared to the standard one, event if the relative VE against the vaccine strains is not as high as expected. This is clearly the result of the additional protection of the adjuvanted vaccine against the second B strain.

If both vaccines protect only against the three vaccine strains they contain (Trivalent scenario in Table 4.2), VE against all influenza cases, as in the original trial protocol, are lower than expected due to the mismatch between the vaccine and the circulating B strains. At the expected VE levels, the probabilities of success in showing an additional gain of the adjuvant is only 48% for a median estimated relative VE of 0.15.

Finally, when both vaccines protect only against the vaccine strains and only matching cases are considered for the analyses (Trivalent Match scenario in Table 4.2), the median estimated relative VE are very close to the simulated values and the probabilities of success are high (82% if all subjects are susceptible). Decrease in relative VE however results in smaller vaccine effects and consequently much lower probabilities of success.

In all cases, the presence of an immune fraction $\pi_0 = 0.2$ leads to a smaller number of cases than expected and thus smaller probabilities of success. When the absolute VE of the standard vaccine is smaller than assumed but the relative VE of the new vaccine is as expected, all other parameters being equal, the number of cases is higher resulting in higher probabilities of success.

VE scenario	Stand. vs. Pbo	Adj. vs. Stand.[1]	π_0	Est. Adj. vs. Stand.[2]	2.5%-97.5% percentiles	P(success)
Mixed	**0.50**	**0.30**	**0**	0.45	0.34, 0.55	1
			0.2	0.45	0.32, 0.56	1
	0.10	**0.30**	**0**	0.31	0.20, 0.41	1
			0.2	0.32	0.19, 0.42	0.99
	0.30	**0.30**	**0**	0.36	0.25, 0.46	1
			0.2	0.37	0.24, 0.47	1
	0.50	**0.25**	**0**	0.41	0.29, 0.52	1
			0.2	0.41	0.28, 0.53	1
	0.50	**0.15**	**0**	0.34	0.20, 0.45	1
			0.2	0.34	0.19, 0.46	0.99
	0.50	**0.05**	**0**	0.26	0.12, 0.38	0.93
			0.2	0.27	0.11, 0.40	0.88
	0.50	**0**	**0**	0.22	0.08, 0.35	0.81
			0.2	0.24	0.07, 0.37	0.74
Trivalent	**0.50**	**0.30**	**0**	0.15	0.00, 0.28	0.48
			0.2	0.15	-0.03, 0.30	0.39
	0.10	**0.30**	**0**	0.20	0.07, 0.30	0.82
			0.2	0.20	0.06, 0.32	0.76
	0.30	**0.30**	**0**	0.17	0.04, 0.29	0.66
			0.2	0.18	0.02, 0.31	0.60
	0.50	**0.25**	**0**	0.12	-0.04, 0.26	0.35
			0.2	0.13	-0.05, 0.28	0.31
	0.50	**0.15**	**0**	0.07	-0.10, 0.21	0.15
			0.2	0.07	-0.11, 0.23	0.13
	0.50	**0.05**	**0**	0.03	-0.14, 0.17	0.06
			0.2	0.03	-0.17, 0.19	0.05
	0.50	**0**	**0**	0	-0.18, 0.15	0.02
			0.2	0.01	-0.19, 0.17	0.02
Trivalent Match	**0.50**	**0.30**	**0**	0.29	0.10, 0.44	0.83
			0.2	0.29	0.07, 0.46	0.68
	0.10	**0.30**	**0**	0.29	0.15, 0.41	0.97
			0.2	0.30	0.14, 0.43	0.93
	0.30	**0.30**	**0**	0.29	0.13, 0.42	0.89
			0.2	0.29	0.11, 0.44	0.84
	0.50	**0.25**	**0**	0.24	0.04, 0.40	0.64
			0.2	0.25	0.02, 0.42	0.55
	0.50	**0.15**	**0**	0.15	-0.07, 0.32	0.27
			0.2	0.15	-0.10, 0.35	0.22
	0.50	**0.05**	**0**	0.04	-0.20, 0.23	0.06
			0.2	0.06	-0.21, 0.27	0.06
	0.50	**0**	**0**	-0.01	-0.26, 0.19	0.02
			0.2	0	-0.27, 0.22	0.02

[1] Additional protection of the adjuvanted vaccine against the vaccine strains.

[2] Estimated additional protection of the adjuvanted vaccine against all strains (mixed and trivalent scenarios) or vaccine strains (trivalent match scenario)

Table 4.2: Estimated VE of the adjuvanted vaccine (Adj.) vs. the standard vaccine (Stand.): median estimates with the 2.5-97.5% inter-percentile range of the estimated values and proportions of simulated trials in which VE is significantly greater than 0, at the 95% confidence level.

While our simulated trials do not perfectly match the characteristics of the original trial, we show that departure from the hypotheses set in the design stage rapidly lead to smaller probabilities of success than expected. Because of the uncontrolled issues that can occur during one influenza season, such as mismatch between vaccine and circulating strains, one has to be cautious when designing a VE clinical trial. The impact of such departures should therefore be thoroughly studied and designs allowing to decrease the risk of failure should be preferred. For example, we recommend that trials be conducted over several seasons. Also, our simulation results weigh in favour of the development of quadrivalent seasonal influenza vaccines which contain both B strains thus limiting the risks of mismatch.

4.4 Discussion

In the epidemiological context, simulation models have been developed in order to study the spread of influenza across networks (Chao et al. 2010). In those models, individuals are related through a social network where influenza spreads, travelling from person to person. In clinical trials however, subjects are not related and the social course of the disease (who catches the disease from whom) is not known. One of our main contributions relies upon our use of historical data in order to get around this unknown information. It should be noted that while our simulation model is inspired by the epidemiological models, we do not aim at capturing all the details of the real situation. Our objective is rather to obtain data consistent with the predominant characteristics of seasonal influenza.

Following Halloran et al. (1996), we define disease occurrence as the combination of the exposure to infection and the susceptibility of the subject. Exposure to infection is relative to the rate of contact, the prevalence of the infection among the population and the infectiousness of the disease. Susceptibility is a function of individual fragility and vaccine protection.

Our simulation method is a flexible and powerful tool. It allows to design and to question the probability of success of seasonal influenza VE trials under varying conditions: design over one vs. several influenza seasons, varying statistical analysis models (Benoit et al. Conditionnaly accepted), good matching season vs. mismatching, homogeneous vs. heterogeneous populations etc. It has the advantage of first not requiring a parametrical shape for the baseline hazard and second of including several sources of heterogeneity. Also, the inclusion of strain specific information allows the assessment of the impact of mismatching on the power of the design.

One possible specific application of our methodology, as illustrated in this chapter, is to retrospectively identify potential sources of problem for a large trial that failed to show significant additional gain of VE for a new adjuvanted vaccine compared to

the standard vaccine. We added an additional placebo group to the trial, which was not ethically acceptable in reality due to the level of fragility of the population of interest. We were able to simulate different scenarios of vaccine protection for the adjuvanted vaccine, both in terms of relative VE but also with different hypotheses of strains protection range. We clearly showed that small departures from the assumed hypotheses could rapidly lead to smaller probabilities of success and that the choice of the primary endpoint (all cases or only matching cases) was crucial.

When designing future trials, our proposed strategy consists of using past data in order to predict a range of possible outcomes. So in order to select an efficient study design, we suggest to run simulations based on data from several historical seasons. By testing the design over both typical virus circulation years and more unusual seasons, such as season 2009-2010, characterized by the H1N1 pandemic, the probability of a successful trial in different contexts can be assessed. Sources of data used to simulate the prevalence of the disease among the studied population can be historical databases but also information from previous clinical trials.

Finally, while we developed this method in the context of seasonal influenza, it could be applied to any other airborne transmission disease such as tuberculosis (Ulrichs 2010) or pertussis (Longini Jr and Halloran 1996) with limited adaptations.

4.5 Conclusion

In this chapter, we developed a methodology based on a simulation model. Our model is a flexible and a powerful tool to design new VE trials. As an illustration, we explored the impact of departures from the protocol hypothesis on the probabilities of success of phase III VE trials. We found that even small departure from the assumptions, especially at the strains protection level, could result in much decreased probabilities of success. Our recommendation is to conduct sensitivity analyses when designing a trial by simulating several scenarios.

In the next chapter, we will use our simulation framework to assess the performances of the classical regression models in estimating seasonal influenza VE.

Summary of Chapter 4

- Symptomatic influenza depends upon exposure to the virus, transmission of the virus, fragility of the host and vaccine protection. Fully parametrical simulation models do not take into account those particularities.
- We developed a flexible model for the simulation of flu VE data, including strains information and historical data.
- We illustrated our methodology by re-simulating a large failed phase III trial and we showed that small departures from protocol assumptions rapidly led to decreased power to show significant VE.
- We recommend to conduct sensitivity analyses when designing an influenza vaccine trial and our simulation algorithm is a powerful tool to accomplish this goal.

Chapter 5

Performances of regression models in estimating seasonal influenza vaccine efficacy

This Chapter is mainly based on the paper "Performances of regression models in estimating communicable infectious diseases vaccine efficacy" by Benoit A., Legrand C. and Dewé W., and submitted to PLOS One.

Vaccine efficacy is classically estimated through logistic, Poisson or Cox regression models. The use of those models is increasingly being challenged in the context of seasonal influenza, where sources of heterogeneity are multiple. Indeed, seasonal patterns, subject heterogeneity, exposure effects and vaccine protection mechanisms conflict with models assumptions and as a result, the estimates of vaccine efficacy could be biased.

In this chapter, we investigate whether this omission of sources of heterogeneity would lead to biased estimates. We use our simulation model in two complementary approaches. First, we derive the marginal vaccine efficacies for different heterogeneity settings. Second, we perform a large simulation study to examine the quality of vaccine efficacy estimates obtained from the regression models when sources of heterogeneity are omitted.

We show that the estimates are indeed biased and that the magnitude of the bias is related to the incidence of the disease. However, at low attack rates, biases are small and not clinically relevant.

5.1 Introduction

To estimate VE against seasonal influenza, clinical trial data are classically analysed using logistic, Poisson or Cox regressions models, adjusting for potential confounders (Nauta 2010; Halloran et al. 1996) and including vaccine group as a fixed effect covariate. However, the assumptions of these regression models are increasingly being challenged in the context of infectious diseases like seasonal influenza (Dewé et al. 2013). As a result, when designing such trials, long discussions between the manufacturers and the regulatory authorities are often engaged to justify the correct use of the selected model. Indeed, VE estimation iscomplicated by the seasonality of the epidemics (Grassly and Fraser 2006), low AR (Atkinson et al. 2011), subject heterogeneity (Struchiner and Halloran 2007; Michiels et al. 2005; McElhaney et al. 2006; Dewé et al. 2013) and variable mechanisms of vaccine protection (Smith et al. 1984). Consequently, the assumptions of the classical statistical models do not always hold and there is concern about biased estimates of VE.

The objective of this Chapter is to assess the quality of the VE estimates from the classical models when these particularities are not taken into account. While this has already partially been discussed (Bretagnolle and Huber-Carol 1988; Lachin 2011), there is clearly a need to assess the magnitude of the biases and their specific impact on seasonal influenza VE trials.

We first briefly describe the classical models for the estimation of VE and the issues risen in the context of seasonal influenza. We then show that time-constant VE (no waning) at the individual level, i.e., conditional VE, is associated with time-dependent population-averaged VE, i.e. marginal VE, when the risk of disease is heterogeneous among the individuals. We next perform a large simulation study where seasonal influenza trials efficacy data are generated through the algorithm presented in Chapter 4 (Benoit et al. Conditionnaly accepted) and analysed with regression models omitting sources of heterogeneity. We also discuss the applicability of more complex models, univariate frailty and cure models and of a more simple model, a fully parametrical survival model assuming constant risk over time. In the last section, we discuss our results and the limitations of our work before suggesting further perspectives.

5.2 Methods of analysis and issues

5.2.1 Classical models

In VE trials for seasonal influenza, the endpoint of interest is the occurrence of well-defined events such as laboratory-confirmed cases of influenza illness. Logistic, Poisson and Cox regressions are the most often used regression models to estimate VE.

These models are presented and discussed in Chapter 2 and a brief overview of general formulation is presented in Chapter 2. Each model is characterized by a different approach to modelling of this endpoint, respectively cumulative incidence, incidence taking into account the time of exposure, and time-to-event.

5.2.2 Issues with the classical models

Logistic, Poisson and Cox regressions all assume a multiplicative effect of vaccine on the baseline risk of disease. Time of exposure is not taken into account in the logistic model. In VE trials however, time of exposure is usually very similar between subjects as accrual is done over a very short period and observation times are usually short (e.g. one season), leading to very low drop-out rates. The three models are thus expected to perform similarly with regards to exposure.

Seasonal influenza is characterized by a periodic surge in disease incidence (Grassly and Fraser 2006). Those seasonal variations may be an issue for the logistic and the Poisson regression models as they assume a constant rate of event within the period of observation. However, it does not affect the semi-parametric Cox regression model which allows a time-varying baseline risk when all subjects enter the observation period at the same time.

An important characteristic of seasonal influenza is a low annual incidence rate (Atkinson et al. 2011). By the end of the annual season, less than 5% of the population have been infected. In the specific case of constant baseline risk, short follow-up times, low incidences and small RR associated with the risk factors the three regression models have been shown to lead to similar estimates of VE (Frome 1983; Callas et al. 1998; McNutt et al. 2003). When the outcome is more frequent, the OR is no longer a good approximation of the RR. In this situation, erroneously relying on the OR winds up overestimating the association between the covariate and the outcome (Zhang and Kai 1998).

The three models make the assumption that, given the covariates, observations are independent and identically distributed. In the context of vaccination, unobserved heterogeneity sources may cause the subjects to have various levels of fragility to the disease. When inter-subjects heterogeneity exists, this assumption is not verified and biased VE estimates may be obtained (Wienke 2010).

Regarding the mode of action of the vaccine, as mentioned in Chapter 2, there is most probably a mixture of partial protection and complete protection . The population under study thus combines a totally immune and a partially immune sub-populations. None of the studied models is adapted in this situation (Halloran et al. 1992). One solution would be to combine the logistic regression with a model accounting for the time of exposure. The mixture cure model (Boag 1949; Berkson and Gage 1952) makes such a combination. However, it requires sufficient follow-up to differentiate

between the non-susceptible fraction and the susceptible who have not been exposed enough. In clinical trials for VE against seasonal influenza, the follow-up is too short to make this distinction.

5.3 Conditional versus marginal VE

Conditional VE is the efficacy at the individual level, i.e. conditional on the individual characteristics. Marginal VE is the averaged VE at the population level. In this section, we show that time-constant individual VE is associated with time-dependent population-averaged VE when the risk of disease is heterogeneous among the individuals.

In Tournoud and Ecochard (2008), the authors develop alternative distributions for the promotion time cure model. They derive marginal survival function for four distributions including Bernoulli and the negative binomial. In order to derive the marginal time-dependent VE for different types of heterogeneities, we follow their methodology by considering the general model:

$$h_{cond}(t|\psi) = \psi h_{cond}(t) \tag{5.3.1}$$

where $h_{cond}(t|\psi)$ is the conditional hazard function of developing influenza at time t and ψ is a latent heterogeneity variable. The marginal survival function of developing a influenza is:

$$\begin{aligned} S_{pop}(t) &= E_{\psi}\left(S_{cond}(t|\psi)\right) \\ &= \int_0^{\infty} S_{cond}(t|u) f_{\psi}(u) du \\ &= \int_0^{\infty} \exp\left(-H_{cond}(t|u)\right) f_{\psi} du \\ &= \int_0^{\infty} \exp\left(-H_{cond}(t)u\right) f_{\psi} du \\ &= \mathcal{L}_{\psi}(H_{cond}(t)) \end{aligned} \tag{5.3.2}$$

where $\mathcal{L}_{\psi}$ denotes the Laplace transform of ψ (Tournoud and Ecochard 2008).

We apply this transformation to the model developed in Chapter 4 for 3 mechanisms of heterogeneity at the individual level: a strictly positive continuous frailty term, z, a fraction of non-susceptible subjects, π and a positive discrete fragility term c_i. We combines them in an individual time-to-event data generation model. We define a

mixture cure model based on the principles that to get symptomatic influenza a subject must be in contact with an infected person, must be susceptible and transmission must occur (Halloran et al. 1996; Benoit et al. Conditionnaly accepted). In our model, z_i represent the susceptibility of a subjects for example through his/her immunological profile, π_0 and π_1 are respectively the proportions of immune subjects in the reference and the experimental groups and c_i is the number of daily contacts that are relevant to influenza transmission. To simplify the notation we consider a $1:1$ ratio and, for each vaccine group, i goes from 1 to $\frac{n}{2}$.

$$\begin{aligned} h_{i,0}(t) &= (1-\omega_{i,0})\{z_i c_i \rho\, p(t)\} \\ h_{i,1}(t) &= (1-\omega_{i,1})\{(1-VE_S)\, z_i c_i \rho\, p(t)\} \end{aligned} \tag{5.3.3}$$

Where $i = 1, ..., \frac{n}{2}$ is the subject indicator, $g = 0, 1$ is vaccine group indicator and $h_{0,i}(t)$ and $h_{1,i}(t)$ are the hazard functions in the reference group and in the experimental vaccine group respectively.

We consider two mechanisms of VE: complete protection for a proportion of the subjects (π_0 and π_1) and partial protection for the susceptibles (VE_S). Total efficacy level is computed as $1 - \left\{\frac{1-\pi_1}{1-\pi_0}(1-VE_S)\right\}$ (Halloran et al. 1992). The susceptibility status, $\omega_{i,0}$ or $\omega_{i,1}$, follows a Bernoulli distribution with respectively parameter π_0 or π_1 and is equal to 0 when the subject is susceptible to the disease, 1 otherwise. We consider a number of daily contacts relevant to the spread of infection for each subject. The contact term c_i is drawn from a positive distribution and bounds the cumulative hazard of the survival part. The probability that a contact is infectious at time t is introduced through the prevalence of the disease, $p(t)$. Fragility is included as a frailty term z_i (Duchateau and Janssen 2007). Finally, the hazard is proportional to the the transmission probability ρ of the disease.

5.3.1 Individual frailty

In our conditional model, the individual instantaneous risk of disease is multiplied by a fragility level z_i drawn from a strictly positive continuous distribution. Since the shape of the frailty distribution does not impact the estimation of the fixed effects (Pickles and Crouchley 1995) and following Duchateau and Janssen (2007), we propose to use a one-parameter gamma distribution with variance parameter δ with $\delta > 0$. We write the Laplace transform

$$\mathcal{L}_{\text{ga}}(s) = (1+\delta s)^{-\frac{1}{\delta}} \tag{5.3.4}$$

and the corresponding marginal survival function is then

$$\begin{aligned} S_{marginal}(t) &= \mathcal{L}\left(-\log(S_{cond}(t))\right. \\ &= \left(1-\delta\log\left(S_{cond}(t)\right)\right)^{-\frac{1}{\delta}} \end{aligned} \tag{5.3.5}$$

where S_{cond} corresponds to the conditional survival function for $z = 1$. The corresponding marginal hazard is

$$h_{marginal}(t) = \left(1-\delta\log\left(S_{cond}(t)\right)\right)^{-1} h_{cond}(t) \tag{5.3.6}$$

where h_{cond} is the conditional hazard function.

We now compare the conditional ($HR_{cond} = \exp(\beta_1)$) and the marginal hazard ratio for the vaccine effect binary covariate and we derive the marginal HR

$$\begin{aligned} HR_{marginal} &= \frac{h_{marginal,1}(t)}{h_{marginal,0}(t)} \\ &= \frac{\left(1-\delta\log\left(S_0(t)\right)\exp(\beta_1)\right)^{-1} h_0(t)\exp(\beta_1)}{\left(1-\delta\log\left(S_0(t)\right)\right)^{-1} h_0(t)} \end{aligned} \tag{5.3.7}$$

and the marginal VE

$$\begin{aligned} VE_{marginal}(t) &= 1 - HR_{marginal} \\ &= 1 - \left\{\frac{\left(1-\delta\log\left(S_0(t)\right)\right)}{\left(1-\delta\log\left(S_0(t)\right)\exp(\beta_1)\right)}\exp(\beta_1)\right\} \end{aligned} \tag{5.3.8}$$

where $S_0(t)$ is the survival function at time t in the reference group conditional on the frailty and $exp(\beta_1)$ is the vaccine effect.

Marginal VE decrease with AR and the departure is directly related to the conditional VE level and the variability of the frailty term.

Figure 5.1 A shows VE as a function of AR in the reference group for 4 levels of leaky vaccine efficacy (0, 0.3, 0.5 and 0.7) and $\delta = 4$. This value leads to a large heterogeneity in the levels of fragility at the beginning of the observation period, based on the fact that post-vaccination immunogenicity markers, usually the disease antibody titres, tend to be very heterogeneous among subjects having received the same vaccine (Thomas and Moridani 2010). Even in presence of a highly variable fragility term, when the outcome of interest is uncommon (below 10%) and thus the cumula-

tive hazard is low, only small departures from the initial VE are observed. The size of the departure is positively related to the level of heterogeneity.

5.3.2 Number of daily contacts

At the individual level, contacts rate heterogeneity is introduced through a multiplicative term with a positive discrete distribution. Following Mossong et al. (2008) we draw the contact term from a negative binomial distribution with parameters ϕ and η, with $\phi > 0$ and $\eta \geq 0$. Following Tournoud and Ecochard (2008), we write the Laplace transform

$$\mathcal{L}_{\text{nb}}(s) = \left(\frac{\phi}{\phi + \eta(1 - \exp(-s))} \right)^{\phi} \tag{5.3.9}$$

The corresponding survival function is then

$$\begin{aligned} S_{marginal}(t) &= \mathcal{L}\left(-\log(S_{cond}(t))\right) \\ &= \exp\left\{ -\phi \log\left[1 + \frac{\eta}{\phi}\left(1 - S_{cond}(t)\right) \right] \right\} \end{aligned} \tag{5.3.10}$$

where S_{cond} corresponds to the conditional survival function for $c = 1$. The corresponding marginal hazard is

$$h_{marginal}(t) = \frac{\eta S_{cond}(t) h_{cond}(t)}{\left[1 + \frac{\eta}{\phi}\left(1 - S_c(t)\right) \right]} \tag{5.3.11}$$

where h_{cond} is the conditional hazard function.

We now compare the conditional ($HR_{cond} = \exp(\beta_1)$) and the marginal hazard ratios for the vaccine effect binary covariate and we derive the marginal HR

$$\begin{aligned} HR_{marginal} &= \frac{h_{marginal,1}(t)}{h_{marginal,0}(t)} \\ &= \exp(\beta_1) \frac{\phi + \eta\left(1 - S_0(t)\right)}{\phi + \eta\left(1 - S_0(t)\right)^{\exp(\beta_1)}} S_0(t)^{\exp(\beta_1) - 1} \end{aligned} \tag{5.3.12}$$

and the marginal VE

$$\begin{aligned} VE_{marginal}(t) &= 1 - \frac{h_{marginal,1}(t)}{h_{marginal,0}(t)} \\ &= 1 - \left\{ \exp(\beta_1) \frac{\phi + \eta\left[1 - S_0(t)\right]}{\phi + \eta\left[1 - (S_0(t))^{\exp(\beta_1)}\right]} S_0(t)^{\exp(\beta_1)-1} \right\} \end{aligned} \tag{5.3.13}$$

Marginal VE decreases with AR. The departure is positively related to the conditional VE level and to the two parameters of the negative binomial contact term.

Figure 5.1 B shows VE as a function of AR in the reference group and for 4 levels of leaky vaccine efficacy (0, 0.3, 0.5 and 0.7), for a negative binomial contact term characterized by $\eta = 18$ and $\phi = \frac{1}{0.36}$ (Mossong et al. 2008). Marginal VE quickly decrease over time and proportionally to the initial VE.

5.3.3 Vaccine mechanism of protection

When the mechanism of vaccine protection is a mixture of an all-or-none and a leaky mechanisms, the studied population will combine both a totally and a partially immune subpopulations of individuals. To consider both efficacy mechanisms, we can use a mixture model where the overall survival of the population, $S_{marginal}(t)$, is decomposed into two components: a proportion π (corresponding to π_0 or π_1 depending on the vaccine group) of immune (not susceptible) subjects and a proportion $1 - \pi$ (corresponding to $1 - \pi_0$ or $1 - \pi_1$ depending on the vaccine group) of the population whose time-to-event are defined by a survival function $S(t)$. The two components of the model are affected by the vaccine.

Following Hougaard (1999), we propose to use a Bernoulli distribution with parameter π for the proportion of immune subjects with $0 \leq \pi \leq 1$. Following Tournoud and Ecochard (2008), we write the Laplace

$$\mathcal{L}_{\mathcal{B}e}(s) = (1 - \pi) + \pi \exp(-s) \tag{5.3.14}$$

and the corresponding marginal survival function is then

$$\begin{aligned} S_{marginal} &= \mathcal{L}\left(-\log(S_{cond}(t))\right) \\ &= \pi + (1 - \pi) S_{cond}(t) \end{aligned} \tag{5.3.15}$$

where S_{cond} corresponds to the conditional survival function, i.e. the survival function of the susceptible fraction of the population. The corresponding marginal hazard is

$$h_{marginal}(t) = \frac{(1-\pi)h_{cond}(t)S_{cond}(t)}{\pi + (1-\pi)S_{cond}(t)} \tag{5.3.16}$$

where h_{cond} is the conditional hazard function.

We now compare the conditional ($HR_c = \exp(\beta_1)$) and the marginal hazard ratios for the vaccine effect binary covariate and we derive the marginal VE.

$$\begin{aligned} HR_{marginal} &= \frac{h_{marginal,1}(t)}{h_{marginal,0}(t)} \\ &= \exp(\beta_1)\frac{1-\pi_1}{1-\pi_0}\frac{\pi_0 + (1-\pi_0)S_0(t)}{\pi_1 + (1-\pi_1)S_0(t)^{exp(\beta_1)}}S_0(t)^{\exp(\beta_1)-1} \end{aligned} \tag{5.3.17}$$

$$\begin{aligned} VE_{marginal}(t) &= 1 - \frac{h_{marginal,1}(t)}{h_{marginal,0}(t)} \\ &= 1 - \left\{\exp(\beta_1)\frac{1-\pi_1}{1-\pi_0}\frac{\pi_0 + (1-\pi_0)S_0(t)}{\pi_1 + (1-\pi_1)S_0(t)^{exp(\beta_1)}}S_0(t)^{\exp(\beta_1)-1}\right\} \end{aligned} \tag{5.3.18}$$

where $exp(\beta_1)$ is the vaccine effect in the survival portion of the model, π is the proportion of subjects totally protected by the vaccine and $S_0(t)$ is the survival function at time t in the reference group.

Marginal VE tends to increase with AR and with the relative proportion of the type II (all-or-none) mechanism in the total conditional VE.

As an illustration, we have computed VE as a function of AR for a mixed protection VE mechanism with $\pi_0 = 0$ (all subjects from the reference group are susceptible) and $\pi_1 = 0.20$ and total efficacy of 0, 0.3, 0.5 and 0.7 (Figure 5.1 C).

5.4 Simulations study

When estimating VE without accounting for subject heterogeneity, we make the assumption that VE is constant over time. However, we showed that the inclusion of sources of heterogeneity resulted in time-varying marginal VE. Consequently, if the

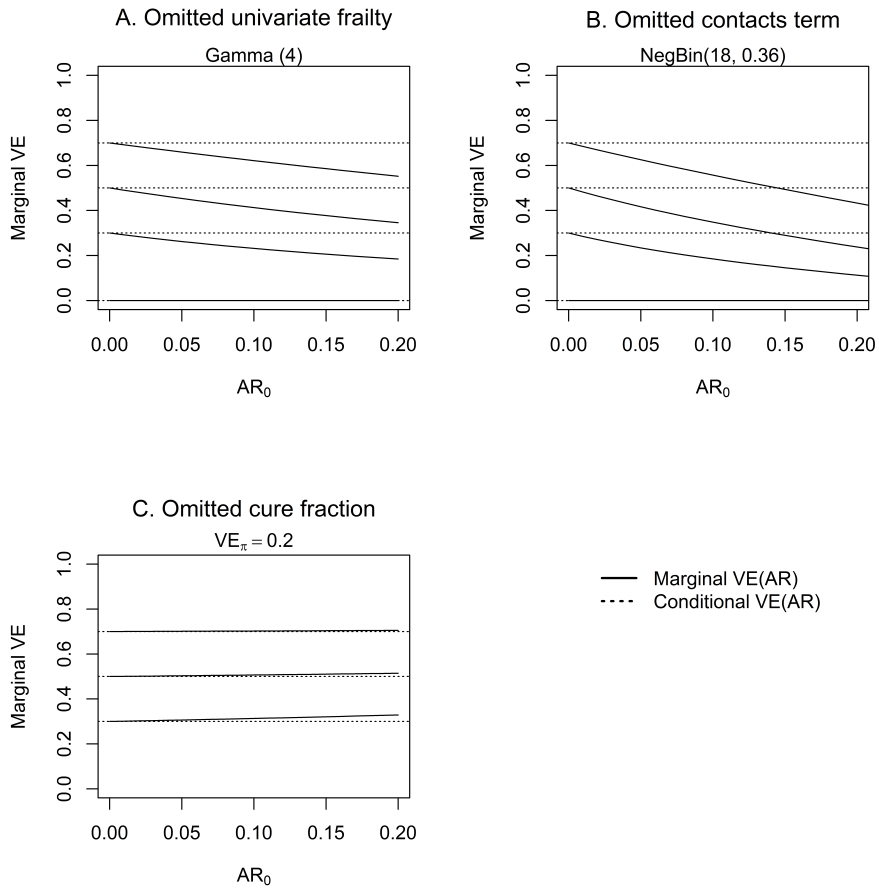

Figure 5.1: Marginal VE as a function of conditional VE and AR in the susceptible reference group.

heterogeneity is not accounted for in the analysis model, the estimated VE will actually correspond to an average efficacy over the period of observation.

5.4.1 Simulation setting

We run a large simulation study to quantify the amount of the biases resulting from misspecified regression models. We use the conditional model described in Chapter 4 to generate clinical trial data. Because we want to include seasonal effect in our generated data, we derive the probability that a contact is infected, $p(t)$, from historical data (Benoit et al. Conditionnaly accepted).

For each of the 3 scenarios described above and an additional one combining all heterogeneity sources, 500 trials with 10000 subjects equally allocated to the two treatment groups and two countries were simulated. Prevalences $p(t)$ were derived from FluNet (Flahault et al. 1998) surveillance data for the German and the U.S. flu season 2000-01 for all circulating influenza strains. Our data generation algorithm allows to capture the variations in occurrence time, magnitude of the epidemic peaks as well as the strains repartition over time and geographical regions (Benoit et al. Conditionnaly accepted). Transmission probability was modulated in order to reach attack rates in the susceptible sub-population within the reference group (AR_0) between 0.02 and 0.20.

VE was estimated through logistic, Poisson, Cox regression models including vaccine and country effects. We only considered censoring due to the termination of the trial. Median estimated VE with 95% empirical confidence intervals (percentiles 2.5 and 97.5) are presented as a function of AR_0 and by conditional efficacy levels (Figure 5.2)

5.4.2 Results

When no source of heterogeneity is considered in data generation and a leaky VE mechanism, estimations from the logistic regression are characterized by a slight VE overestimation trend increasing with AR_0. Estimates from the Poisson and the Cox regression models are in close agreement with the true values.

When data are generated with sources of heterogeneity, the logistic regression for low AR, the Cox and the Poisson regression lead to very similar results. Therefore, we only present the Cox regression estimates.

We first simulate data allowing different levels of fragility of the people under study. Our frailty term is generated from a one-parameter gamma distribution with $\delta = 4$ as in the previous section. In this case, VE from the misspecified Cox model slightly un-

derestimates the true efficacy. Underestimations are more important at higher efficacy levels and for higher baseline attack rates.

VE estimated from a misspecified Cox regression on data simulated with heterogeneity in the daily number of contacts gives estimates very close to the real efficacy levels used in the simulations, although a slight trend toward underestimations can be seen as AR_0 increases.

The misspecified Cox regression very slightly overestimates VE when the vaccine has a mixed mechanism of protection. Overestimations are visible at higher baseline attack rates and when the all-or-none part of the protection mechanism is more prevalent, at smaller VE levels.

When the studied sources of heterogeneity are combined, VE estimates from a misspecified Cox regression tend to be smaller than the simulated efficacy levels. The bias increases both VE and baseline attack rates.

At low event rates, all models gave good VE estimates, even when sources of heterogeneity are omitted in the analysis model. Based on our results, it seems that, whenever AR are low (as is expected for many infectious diseases), models taking into account the time of exposure, such as the Poisson or the Cox model should be preferred. More complex models should be considered only in case of larger AR.

5.5 Additional results

In the previous section, we have shown that analysis regression models not taking into account the sources of heterogeneity gave surprisingly good estimates of the true VE at low AR. However, if the source of heterogeneity is of particular interest or when the expected AR are higher, one should use a more complex model. In this section, we first discuss and present simulations results of Cox survival models including an univariate frailty or a cure fraction. On the opposite, we explore the possibility of estimating VE with an even more simple model, the exponential survival model. We show that for low AR, a simple fully parametrical survival model assuming constant hazard gives estimated VE close to the true VE values.

5.5.1 Univariate frailty

We have shown that failing to account for subject heterogeneity in the analysis of VE data did not lead to a clinically-relevant bias when AR were low. However, if we want to model this heterogeneity (either because of a particular interest or in the setting of higher AR), a frailty model can be used (Duchateau and Janssen 2007). In this section,

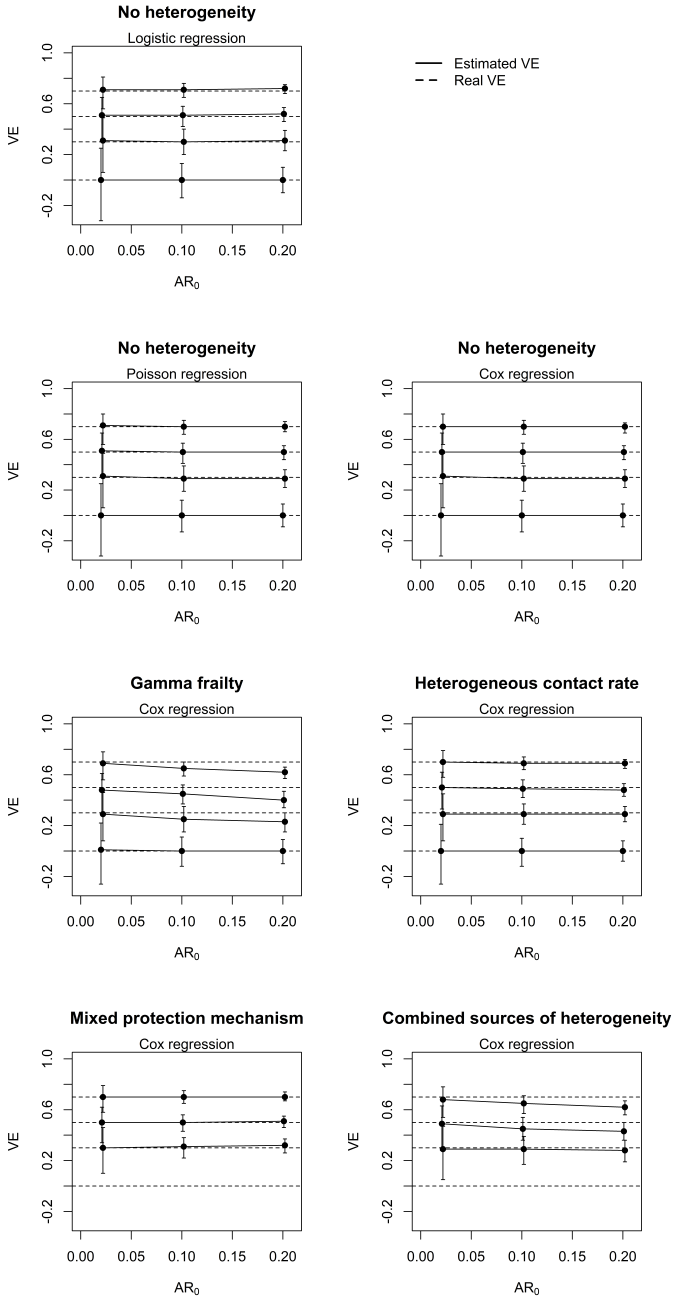

Figure 5.2: Estimated VE (median and estimation 95% CI) over AR_0 by simulated VE and omitted sources of heterogeneity. Dotted lines represent the true VE, solid lines represent the median estimated VE with a 95% CI.

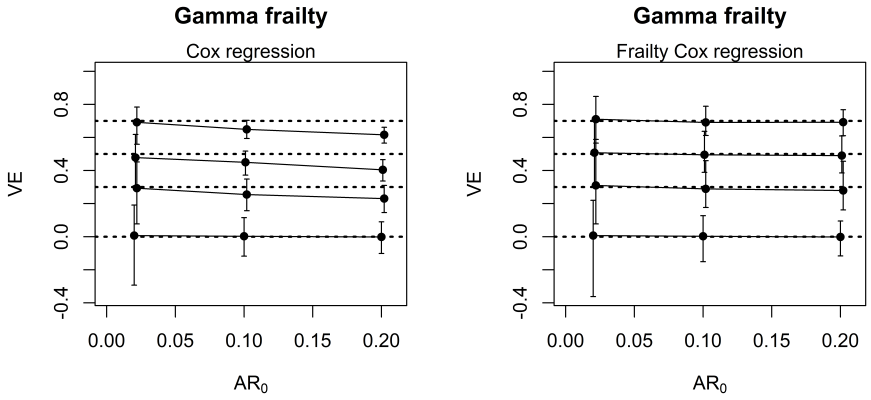

Figure 5.3: VE estimates (with 95% empirical CI) obtained with a Cox regression model univariate frailty (right) or not (left)

we compare the VE estimates from a Cox PH model model accounting for individual heterogeneity and a a Cox PH model including an univariate frailty.

Univariate frailty models (Wienke 2010) allow to account for unobserved individual heterogeneities. The individual hazard function with frailty z for a PH survival model is given in equation 4.2.1 of chapter 4.

We generate data from 500 trials with sample size 10000 including an individual random fragility generated from a one-parameter gamma distribution with $\delta = 4$, in order to reflect a 10-fold risk difference between the mean and the percentile 99 subjects, translating large fragility differences between healthy subjects and subjects with concomitant co-morbidities. We analyse the data using a Cox regression model that either includes or does not include a univariate gamma frailty term. The results of these analyses are presented in figure 5.3.

While including an univariate frailty in the analysis modelled for VE trials data improves the quality of the estimates at higher AR, the assumptions made for this model can be quite strong. The biggest problem is that in a Cox regression model, unobserved heterogeneity cannot be distinguished from a situation of homogeneity under a non-proportional hazards model (O'Quigley and Stare 2002). As a result, when choosing to include an univariate frailty, we make the strong assumption that the underlying cause of non-proportionality is heterogeneity and not time-varying covariate(s) effect(s), for example waning of the vaccine effect over time. Also, a distribution has to be specified for the frailty part of the model. While it has been shown that the choice

of the frailty distribution does not affect the estimation of the survival function parameters (Pickles and Crouchley 1995), misspecification of the frailty distribution might be problematic if one is interested in the random effect.

5.5.2 Cure model

The mixture cure model has been developed by Boag (1949) and Berkson and Gage (1952) in the context of oncology studies.. The overall survival of the population includes two components: a proportion π of cure subjects and a proportion $(1-\pi)$ of the population whose time-to-event can be characterized by a survival function. In the context of seasonal influenza VE, the cured fraction correspond to the subjects who are not susceptible to the disease.

Mixture cure models allow to simultaneously estimate the incidence, i.e. the probability that an event of interest will occur and the latency, i.e. when it will occur given that it occurs (Corbière and Joly 2007). Let ω be the indicator of susceptibility, set to 1 when the subjects is non susceptible, 0 otherwise. The unconditional survival function is given by

$$S(t|X_1 = x_1) = \pi(X_1 = x_1) + ((1 - \pi(X_1 = x_1))S(t|X_1 = x_1, \omega = 0)) \quad (5.5.1)$$

where $S(t|X, \omega = 0)$ is the survival function for the susceptible subjects and $\pi(X_1 = x_1)$ is the probability of being non susceptible for a subject with vaccine group factor x_1. This probability is modelled through of binary regression model (Corbière and Joly 2007). One of the popular models is the logistic regression model where

$$\text{logit}(\pi(X_1 = x_1)) = \alpha_0 + \alpha_1 x_1 \quad (5.5.2)$$

where α_0 is the intercept and α_1 the vaccine effect parameter associated with the incidence part of the model. The survival function for the susceptible subjects is model through a Cox regression model with β_1 the vaccine effect parameter associated with the latency part of the model.

Total VE (Halloran et al. 1996) is computed as

$$VE_{tot} = 1 - (\exp(\alpha_1) \times \exp(\beta_1)) \quad (5.5.3)$$

The main issue with cure models is how to identify the subjects who are completely protected. Indeed, when the time of follow-up is too short, making the difference between the non susceptible subjects and the subjects who will develop the infection

after the closing of the study is impossible (Taylor 1995; Li et al. 2001). Simplification assumptions have to be made. We apply the frequently used zero-tail assumption that assumes that all non infected subjects by the end of the trial were not susceptible. As such, we wrongly consider that all the non infected subjects were not susceptible.

We model the effect of the vaccine on the cured fraction through a logistic regression model and the survival part of the model through a Cox regression model (Corbière et al. 2009). We fit our model through the SAS macro PPSMCM developed by Corbière and Joly (2007).

We generate data from 250 trials with sample size 10000 and include 20% of non susceptible subjects in the group who received the experimental vaccine. Overall simulated efficacy, including a leaky and a all-or-none portions, is 70% and we generate data with AR levels of 2% and 15%. While the latter level of AR is highly improbable in the context of influenza, it was selected for the purpose of illustration. We analyse the trials data with a logistic regression, a Cox regression and a semi-parametric mixture cure model.

The results of our simulations are presented in figures 5.4 and 5.5. Note that Figures 5.4 and 5.5 are presented in different y-scales in order to preserve their readability. For both AR values, VE estimates from the cure model are more variable than those from the Cox and the logistic regression models. When the AR is greater, the estimates from the cure model suffer from the same drawback as the estimates from the logistic regression model, i.e. the OR are no longer a good approximation of the RR and VE are overestimated.

5.5.3 Exponential model

Throughout this work, we have shown that in the situation where censoring is high, incomplete parsimonious (few parameters) models tend to give good estimates of VE, sometimes even better than more complex models. Because seasonal influenza tends to have non-constant incidences over time, our recommendation is to apply a model that does not make any assumption on the baseline hazard. However, since the Poisson regression gave VE estimates very close to the true values, we decided to explore the quality of the estimates obtained through a the exponential survival model, the simplest possible survival distribution, which assumes a constant risk over time. The hazard is defined as $h(t) = h$.

We generate data from 500 trials with sample size 10000 and including no sources of heterogeneity. We analyse the data with a Cox regression model and with an exponential survival model. The results of these analyses are presented in Figure 5.7 A (Cox regression model) and B (Exponential survival regression model). For illustra-

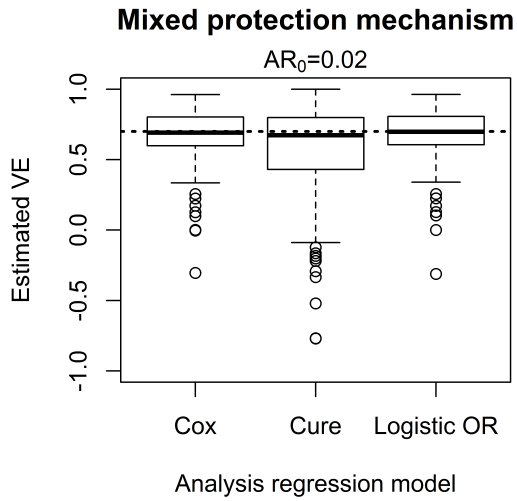

Figure 5.4: VE estimates for a Cox, a cure and a logistic model for data generated with a mixed vaccine protection mechanism for an AR of 2%.

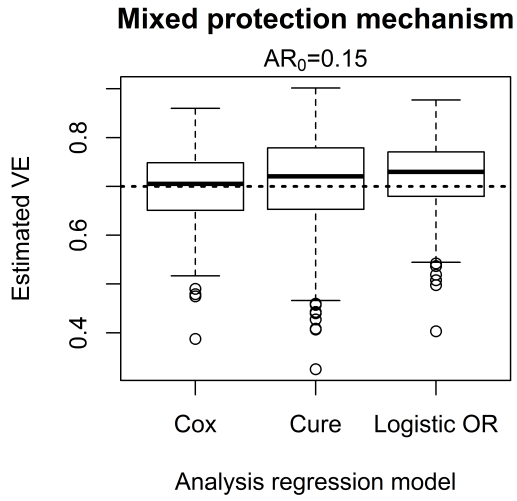

Figure 5.5: VE estimates for a Cox, a cure and a logistic model for data generated with a mixed vaccine protection mechanism for an AR of 15%.

tion purpose, we simulate AR from 1% to 70%. Figure 5.6 shows the historical data used to compute the daily probabilities of a contact infectiousness, $p_k(t)$, where k is the strain indicator.

At low AR, the exponential model give results very similar to the Cox regression model.

5.6 Discussion

It seems from a review performed in Chapter 2 that very simple models are used to analyse Phase III VE trials. However, seasonal influenza epidemics are complex and the use of those models is increasingly being questioned by the regulatory agencies. In this chapter, we assess the quality of the estimates of VE obtained with the logistic, Poisson and Cox regression models in the context of seasonal influenza VE trials.

In presence of heterogeneity and constant conditional VE over time, marginal VE varies with time of exposure. Consequently, marginally, the hazards are no longer proportional. The departure from the initial VE increases with AR_0 and VE levels. At low attack rates however, these departures are small. When analysing data without taking into account heterogeneity, VE is assumed to be constant and the estimated VE correspond to an average efficacy over the period of observation. We ran a simulation study in the context of influenza to quantify the magnitude of the departures that could be obtained.

VE estimates from logistic, Poisson and Cox regression models were found to be unbiased for highly censored seasonal homogeneous data. It would be interesting to extend our work to the analysis of trials data run over several seasons and including a succession of periods with very low or null disease incidence and periods of high disease circulation. With higher baseline attack rates, logistic regression vaccine efficacy estimates based on the odds ratios gave slight overestimations of the true efficacy, as expected (Lachin 2011; Zhang and Kai 1998).

Inclusion of a non susceptible sub-population in the studys led to an overestimation of VE while omission of heterogeneity resulted in an underestimated VE. The results of the simulations are consistent with our analytical results, except in the case of the contacts heterogeneity. In all cases, we showed that at high censoring levels, i.e. low cumulative attack rates over the observation period, the bias was negligible and did not affect the conclusions of the trial.

The effect of combining several sources of heterogeneity is more difficult to predict based on the analytical results. Indeed, departures from the conditional VE go into opposite directions and have varying intensities. We therefore studied this case through

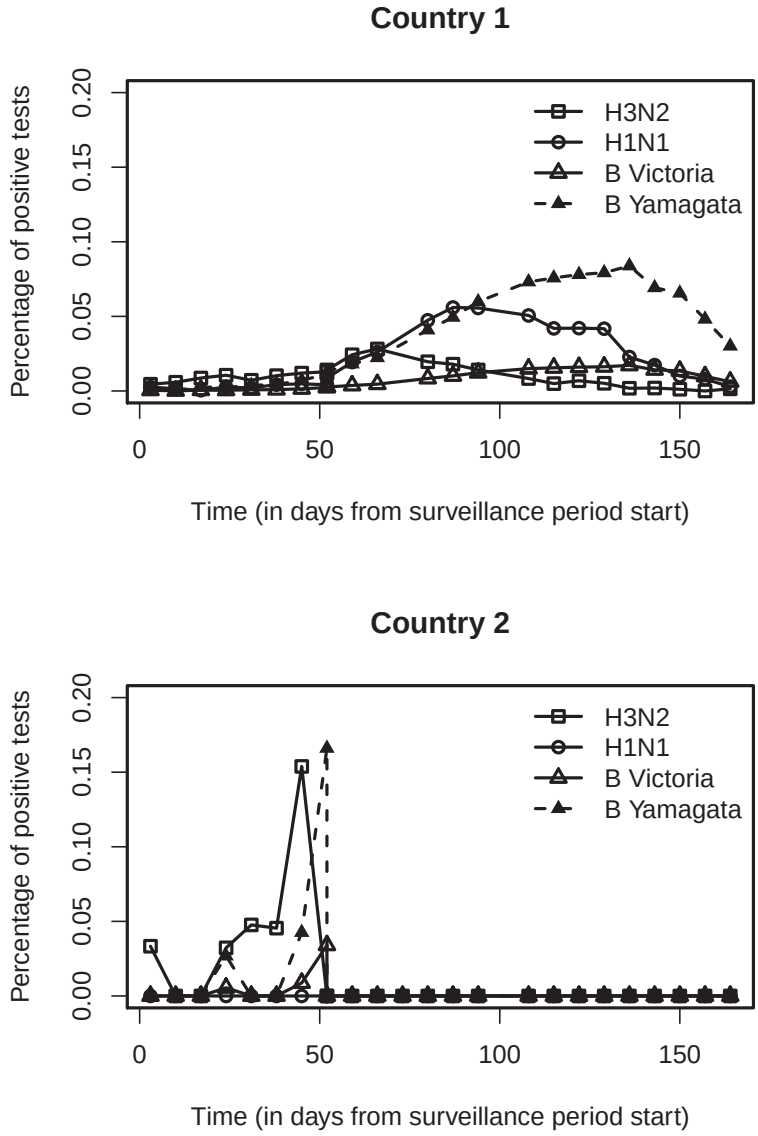

Figure 5.6: Incidences over time of strains of influenza by country for season 2000-2001

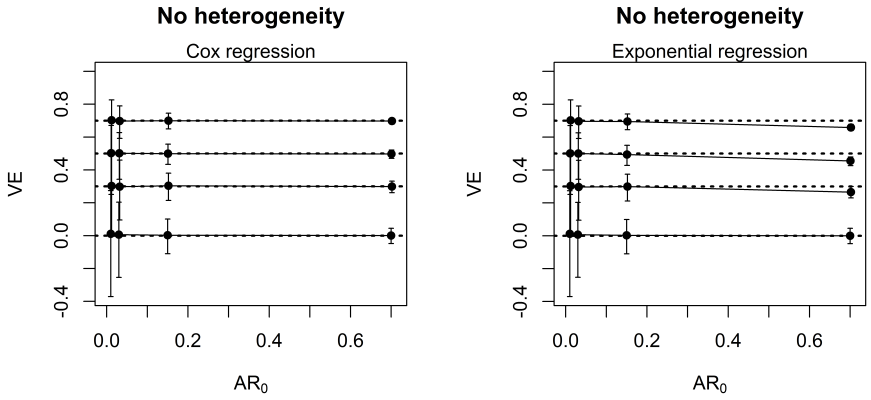

Figure 5.7: VE estimates and 95% CI from a Cox regression model (left) and a fully parametrical exponential survival regression model (right)

simulation. While VE estimates were biased (underestimated), especially with higher AR, in situations similar to the typical influenza season, biases were negligible, with no medically-relevant impact on the conclusions of the trials. Moreover, the estimates tend to be conservative. Poisson and Cox regressions should be preferred to logistic regression that can give slightly biased estimations even if correctly specified.

In the case of higher AR (over 10%), more appropriate models, such as Cox regression including a frailty term, a cure fraction, or a combination of the two, should be preferred. In the case of the cure model, RR should be used instead of OR for the incidence part of the model. These models however have to be interpreted carefully. First, it is impossible to distinguish the effects of heterogeneity from the time-varying covariate effects. For example, the inclusion of a univariate frailty might hide the presence of vaccine waning (Moghadas 2004) or virus mutations over time (Antia et al. 2003). Second, accurately estimating the cured (non susceptible) sub-population requires sufficient follow-up in order to distinguish between the "cured" subjects (Halloran et al. 1996; Li et al. 2001; Peng and Zhang 2008) and the susceptible subjects who have not been infected yet.

In the next chapter, we challenge the traditional paradigm of phase III trials for VE against seasonal influenza. We argue that the actual methodology provides an incomplete answer to the question of VE in the future.

Summary of Chapter 5

- Vaccine efficacy clinical trials data are usually analysed through Cox, logistic or Poisson regression models. However, the particularities of influenza are in contradiction with some assumptions of those models.
- When there is latent heterogeneity at the individual level, marginal vaccine efficacy over time is no longer constant.
- When the model used to estimate vaccine efficacy does not take into account individual heterogeneity the estimated vaccine efficacy is an averaged efficacy over the surveillance period.
- At low attack rates however, departures from the conditional vaccine efficacy are not clinically relevant. Therefore, we recommend to keep using parsimonious data analysis models such as the Poisson and the Cox regression models.

Chapter 6

Accounting for strain heterogeneity in the estimation of vaccine efficacy against seasonal influenza

In this chapter, we propose a new trial design and analysis methodology that allows the characterization of the VE heterogeneity between countries and influenza seasons (referred as clusters). In addition, we also suggest the use of predictive intervals for future cluster-specific VE instead of intervals for the mean VE.

We analyse trial data with a hierarchical survival model taking into account risk heterogeneity and VE heterogeneity across clusters. Our model parameters and the statistical intervals for the cluster-specific VE are estimated using Bayesian statistics.

We illustrate our methodology through a large simulation study and by re-analysing the trial Influence 65. We argue that our methodology provides better insight on the range of future VE across seasons and geographical regions.

6.1 Introduction

Influenza is an infectious disease caused by several virus strains that present varying repartitions between geographical regions and seasons. Typically, a vaccine contains three or four influenza strains and the antigen content is annually reconsidered based on the WHO recommendations. For the same vaccine formulation, pharmaceutical regulations only require efficacy again clinical disease to be shown for a single season, which is performed through a large phase III trial. Subsequent annual modifications of the strain related portion of the vaccine only have to be validated through immunogenicity trials (see Chapter 2 for more details about the development of a vaccine).

Classically, influenza VE trials take place over a single season but over several geographical regions assuming common VE. However, depending on the circulating strains characteristics such as their immunogenicity and their matching levels with the vaccine strains, the vaccinal protection levels may vary from one season/region to another. As a result, the same vaccine can sometimes be proved to be significantly efficacious in one trial but not the other, all other things being equal.

We argue that not accounting for this provides incomplete and unreliable response as for the benefit of the vaccine in the future. We therefore propose to run phase III VE trials over several geographical regions and seasons in order to characterize the VE heterogeneity. We consider VE as the sum of a common quantity to all clusters (season and geographical region) and of a random cluster-specific part. We propose to account for strain heterogeneity between seasons and regions by including estimations of two heterogeneity terms between clusters in the analysis model. Following (Legrand et al. 2009), we suggest the use of a Cox frailty model including two random terms: the random effect at the baseline hazard level accounts for the varying evolutions over time of the seasonal epidemics between clusters while the random effect at the vaccine effect level allows to assess the heterogeneity of VE between clusters, strains heterogeneity. This methodology is already applied in the context of meta-analyses when aggregating individual patient time-to-event data as differences across trials in terms of design, methodology or patients characteristics can contribute to create variability in the treatment effect of interest (Smith et al. 2005; Higgins et al. 2009).

Classically, the primary objective of a CT is to prove that the experimental treatment is likely to have the expected effect in the trial population. To address this objective, confidence intervals and hypothesis tests about the mean treatment effect (FDA 2007) are conducted. The assumption is that, if used in the same population and under the same conditions (compliance and dosage for example), the experimental treatment will have a similar effect once marketed. Yet, in seasonal influenza efficacy trial, because of spacial and temporal heterogeneity, this assumption is less evident. We argue that the objective of a phase III trial should be to prove that the vaccine will

be efficacious in the future. We believe that in the particular context of VE against a heterogeneous disease such as seasonal influenza the decision about VE should not be taken based on a confidence interval for the mean estimated VE, as is actually done, but based on predictive intervals, accounting for the heterogeneity across clusters. Indeed, such information provides insight on the range of future VE across seasons and geographical regions, instead of a mean past season specific vaccine effect. To derive information about VE in the future, i.e. for new clusters, we propose to use predictive intervals instead of confidence intervals. This methodology is routinely used in the industrial context for example where the conformity of new objects manufactured in batches is assessed through tolerance intervals (Wolfinger 1998).

In the Bayesian setting, all model parameters are assigned a probability distribution (Carlin and Louis 2000). Because of this characteristic, the Bayesian algorithm is better suited for solving problems such as computing statistical intervals and especially tolerance intervals (Wolfinger 1998). For this reason, we chose to develop this chapter using Bayesian inference.

In the next section, we propose a short introduction to Bayesian inference. Our suggested analysis model is presented in section 6.3. Types of statistical intervals and their uses are presented in Section 6.4. Implementation of our methodology is detailed in section 6.5. Finally, we illustrate our methodology through a simulation study and a re-analysis of the data from study Influence 65 in Sections 6.6 and 6.7.

6.2 An introduction to Bayesian inference

Gelman (2004) defines Bayesian inference as "the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations". The posterior distribution for θ, the vector of parameters to estimate, is written $p(\theta|y)$ and represents the uncertainty about θ conditionally to the data y (Lunn et al. 2012). Based on Bayes's theorem, it is expressed as:

$$p(\theta|y) \sim L(\theta; y)p(\theta) \tag{6.2.1}$$

where $L(\theta; y)$ is the likelihood of the data given the parameter values and $p(\theta)$ is the prior distribution for θ.

Prior distributions can be vague or informative. The use of vague priors is recommended when one wishes to retrieve all information about the posterior distribution of the parameters of interest from the data in hand while the use of informative priors

acknowledges that there is prior information or intuition on the parameters of interest (Congdon 2007).

To derive the posterior distributions of the parameters of interest, the nuisance parameters have to be integrated out of 6.2.1. Depending on the complexity of the model, this can be done analytically or through numerical solutions (Lunn et al. 2012; Lebrun 2012). When numerical solutions are applied, large samples of instances for the variables to estimate are generated from the posterior joint density and the properties of their distributions are empirically derived from the samples (Lunn et al. 2012). When a full set of conditional posterior distributions can be identified, Gibbs sampler is the most current sampling algorithm: it generates a multi-dimensional Markov chain by splitting the vector of random variables θ into subvectors and sampling each vector, one by one, given the most recent values of all the other elements of θ and the data. When conditional posterior distributions can not be identified, Metropolis sampling algorithms are then used.

6.3 Analysis model

We postulate that estimated VE in a CT arises from the true underlying VE and from virus strains (and sub-strains) related information, such as the matching level between the vaccine and the circulating strains and the strain specific VE. Because strains repartitions change between geographical regions and influenza seasons, further referred as clusters, cluster-specific VE can vary (DiazGranados et al. 2012). One could argue that VE should thus be estimated for each strain. However, because strains composition of the vaccine will change in the future, this would not address the objective of proving that the vaccine will be efficacious for most people who will receive it in the future.

We consider a CT including J clusters ($j = 1, ...J$) each including n_j subjects yielding to a total sample size N. Our analysis model includes two levels of heterogeneity: at the level of the baseline risk and as a random vaccine-effect-by-cluster interaction. We represent the effect underlying the j^{th} of J clusters by $b_{0,j}$. We consider that the $b_{0,j}$ are drawn from a distribution f_{B_0} such that $E[b_0] = 0$ and $var(b_0) = \phi^2$. We represent the interaction effect underlying the j^{th} of J clusters by $b_{1,j}$. We consider that the $b_{1,j}$ are drawn from a distribution f_{B_1} such that $E[b_1] = 0$ and $var(b_1) = \tau^2$.

We define the hazard function for the i^{th} individual in the j^{th} cluster as:

$$\begin{aligned}\lambda_{ij}(t|J=j,X=x) &= \lambda_0(t)\exp(b_{0j}+(\beta_1+b_{ij})x_{ij}) \text{ with}\\ b_{0j} &\sim N(0,\phi^2)\\ b_{1j} &\sim N(0,\tau^2)\end{aligned} \tag{6.3.1}$$

where τ^2 is a measure of the heterogeneity between clusters at the VE level and ϕ^2 of the heterogeneity between clusters at the baseline risk level. We can also write $\lambda_{0j} = \lambda_0(t)\exp(b_{0j})$.

6.4 Statistical intervals

Different types of statistical intervals can be built based on the estimations obtained from our proposed analysis model. The type of interval to be computed will strongly depend on the underlying problem or application (Krishnamoorthy and Mathew 2009).

Intervals are given different interpretations in the Bayesian and the frequentist settings. The frequentist approach regards the model parameters as fixed unknown quantities. In this approach, a statistical interval is interpreted as a range of values that are believed to likely contain the true parameter value (Hahn and Meeker 2011). More precisely, it is interpreted as a range in which the parameter of interest would occur $100(1-\alpha)\%$ of the time with repeated sampling. Bayesians treat model parameters as unknown random quantities and described them probabilistically. They interpret intervals as the region of values, denoted $[L,U]$, that contains $100(1-\alpha)\%$ of the posterior probability of the parameter of interest such that

$$\int_L^U p(\theta|y)d\theta = 1-\alpha \tag{6.4.1}$$

Following this methodology, intervals can be similarly derived for a function of the model parameter(s).

For a parameter θ, the highest posterior density (HPD) interval is the shortest interval containing $100(1-\alpha)\%$ of the posterior sample for θ while the equal-tail is obtained using the empirical percentiles of the posterior distribution of θ. This latter interval definition has been privileged in this chapter.

In this section, we give the definition of each type of interval used in our work. Practical computation of the intervals will later be presented in section 6.5.

Credible intervals

Conventionally, significant VE is statistically showed when the lower bound of the two-sided 95% confidence interval of VE is substantially above zero (CDC 2011) (see Chapter 2 for more details).

A credible interval is the Bayesian analogue of a confidence interval in the frequentist setting.

As shown by Higgins et al. (2009), the presentation of inference only for the mean VE provides an incomplete summary and is highly misleading when there is heterogeneity. In a context such as seasonal influenza where VE includes a cluster-specific component, a credible interval for the trial mean VE does not address the question of the protection of subjects from new clusters (new season and/or new country).

A one-sided interval for the parameter of interest β_1 can be defined as the interval $[L(\beta_1), +\infty[$ (lower bound) or $]-\infty, U(\beta_1)]$ (upper bound) containing $100(1-\alpha)\%$ of the posterior density of β_1. Once the posterior distribution of β_1 is known, the problem reduces to the computation of $L(\beta_1|y)$ or $U(\beta_1|y)$ that satisfies

$$P(\beta_1 \geq L(\beta_1|y)) = 1-\alpha$$

or

$$P(\beta_1 \leq U(\beta_1|y)) = 1-\alpha$$

In the Bayesian setting, a $100(1-\alpha)\%$ credible interval for β_1 is a subset of B_1, where B_1 is the set of possible values for β_1, such that:

$$\int_L^{+\infty} p(\beta_1|y)d\beta_1 = 1-\alpha$$

or

$$\int_{-\infty}^{U} p(\beta_1|y)d\beta_1 = 1-\alpha$$

The posterior density of β_1, denoted $p(\beta_1|y)$ can be derived by averaging the joint posterior density of the model over the nuisance parameters. When this integral has no analytic form, it can be evaluated numerically.

An equal-tail two-sided interval for β_1 can be defined as the interval $[L(\beta_1), U(\beta_1)]$ containing $100(1-\alpha)\%$ of the posterior density of β_1. Once the posterior distribution

of β_1 is known, the problem reduces to the computation of $L(\beta_1|y)$ and $U(\beta_1|y)$ that satisfies

$$P(\beta_1 \geq L(\beta_1|y)) = 1 - \frac{\alpha}{2}$$

and

$$P(\beta_1 \leq U(\beta_1|y)) = 1 - \frac{\alpha}{2}$$

In the Bayesian setting, a $100(1-\alpha)\%$ credible interval for for β_1 is a subset of B_1 such that:

$$\int_L^U p(\beta_1|y)d\beta_1 = 1 - \alpha$$

Prediction intervals

To derive information about a new random effect, we can compute a prediction interval for the random effect of interest, denoted $\beta_{1,new}$. Heterogeneity around the mean effect β_1 is accounted for in the computation of the prediction interval for a new random effect. This interval does also consider the uncertainty of the parameter estimates, that is the imprecision in the estimations of the mean effect β_1 and the variance of the random effect, τ^2. Senn (2004) suggests that such an interval provides a reasonable prediction for a randomly chosen future unit, here a cluster.

We define a new random effect as $\beta_{1,new} = (\beta_1 + b_1|y)$. A one-sided interval for $\beta_{1,new}$ can be defined as the interval $[L(\beta_{1,new}), +\infty[$ (lower bound) or $]-\infty, U(\beta_{1,new})]$ (upper bound) containing $100(1-\alpha)\%$ of the posterior density of $\beta_{1,new}$. Once the posterior distribution of $\beta_{1,new}$ is known, the problem reduces to the computation of $L(\beta_{1,new}|y)$ or $U(\beta_{1,new}|y)$ that satisfies

$$P(\beta_{1,new} \geq L(\beta_{1,new}|y)) = 1 - \alpha$$

or

$$P(\beta_{1,new} \leq U(\beta_{1,new}|y)) = 1 - \alpha$$

A $100(1-\alpha)\%$ prediction interval for $\beta_{1,new}$ is a subset of B_1 such that:

$$\int_{L}^{+\infty} p(\beta_{1,new}|y)d\beta_{1,new} = 1-\alpha$$

or

$$\int_{-\infty}^{U} p(\beta_{1,new}|y)d\beta_{1,new} = 1-\alpha$$

A two-sided prediction interval for a new random effect, $\beta_{1,new}$, can be defined as the interval $[L(\beta_{1,new}), U(\beta_{1,new})]$ containing $100(1-\alpha)\%$ of the posterior density of $\beta_{1,new}$. Once the posterior distribution of $\beta_{1,new}$ is known, the problem reduces to the computation of $L(\beta_{1,new}|y)$ and $U(\beta_{1,new}|y)$ that satisfies

$$P(\beta_{1,new} \geq L(\beta_{1,new}|y)) = 1-\frac{\alpha}{2}$$

and

$$P(\beta_{1,new} \leq U(\beta_{1,new}|y)) = 1-\frac{\alpha}{2}$$

In the Bayesian setting, a $100(1-\alpha)\%$ prediction interval is a subset of B_1 such that:

$$\int_{L}^{U} p(\beta_{1,new}|y)d\beta_{1,new} = 1-\alpha$$

Tolerance intervals

Prediction and tolerance intervals are closely related (Krishnamoorthy and Mathew 2009). Indeed, $100(1-\alpha)\%$ level p-content tolerance intervals are prediction intervals for a fixed proportion p of new random effects. Statistically, a p-content tolerance interval is a prediction interval for specific quantiles of the posterior distribution of a new cluster-specific VE. Analytically deriving these quantities is very complex and numerical solutions are often preferred.

A one-sided p-content tolerance interval is an interval on the distribution of a quantile of the distribution of the random new random effect, $\beta_{1,new}$. We require a $100(1-\alpha)\%$ upper or lower confidence limit for $g_p(\theta)$, the p quantile of $\beta_{1,new}$, based on the posterior distribution $p(g_p(\beta_{1,new})|y)$. Once the posterior distribution of $g_p(\beta_{1,new})$ is known, the problem reduces to the computation of $L(g_p(\beta_{1,new})|y)$ or $U(g_p(\beta_{1,new})|y)$ that satisfies

$$P(g_q(\beta_{1,new}) \geq L(g_q(\beta_{1,new})|y) = 1 - \alpha$$

or

$$P(g_q(\beta_{1,new}) \leq U(g_q(\beta_{1,new})|y) = 1 - \alpha$$

In the Bayesian setting, a $100(1-\alpha)\%$ p-content tolerance interval is a subset of B_1 such that:

$$\int_{L}^{+\infty} p(g_p(\beta_{1,new}|y))dg_p(\beta_{1,new}) = 1 - \alpha$$

and

$$\int_{-\infty}^{U} p(g_p(\beta_{1,new}|y))dg_p(\beta_{1,new}) = 1 - \alpha$$

The computation of a two-sided p-content α-level tolerance interval does not reduce to the combination of separately computed upper and lower limits of two quantiles of the distribution of the new random effect, $\beta_{1,new}$. Except for very simple models, numerical solutions are usually necessary to compute tolerance intervals (Krishnamoorthy and Mathew 2009).

Wolfinger (1998) developed the following methodology: bivariate pairs of $g_{\frac{1-p}{2}}(\beta_{1,new})$ and $g_{\frac{1+p}{2}}(\beta_{1,new})$ are computed from the joint posterior density, forming a sample of size S from the bivariate posterior density of the $\frac{1-p}{2}$ and $\frac{1+p}{2}$ quantiles. A two-sided tolerance interval symmetric about the posterior mean is then built by iteratively searching an interval centered around the posterior mean and containing $100(1-\alpha)\%$ of the $(g_{\frac{1-p}{2}}(\beta^s_{1,new}),\ g_{\frac{1+p}{2}}(\beta^s_{1,new}))$ bivariate pairs, with $s = 1, ..., S$.

6.5 Implementation of the methodology

In this section, we present the practical implementation of our methodology. We first talk about the model parameters estimation. Then, we present a methodology to derive the credible, prediction and p-content tolerance intervals when no analytical solutions exist.

6.5.1 Model estimation

For simplicity and based on results from Chapter 5, we present the case of a completely parametric model in which time-to-event are coming from an exponential distribution, assuming a constant hazard over time, $\lambda_0(t) = \lambda_0 = \exp(\beta_0)$.

To improve the convergence of our model estimation, we apply a hierarchical centering re-parametrization, that is, we assume individual models centered around a population model (Gelfand et al. 1995). b_0 and b_1 are assumed to follow normal distributions centered in β_0 and β_1 respectively. This re-parametrization leads to slight notation changes. From now on, vaccine effect in a new cluster corresponds to $b_{1,new}$. Our model has the form:

$$\lambda_{ij} = \exp((b_{0j}) + (b_{1j}x_{ij})) \tag{6.5.1}$$

With $j = 1, ...J$ and $i = 1, ...n_j$, $b_{0j} \sim N(\beta_0, \phi^2)$ and $b_{1j} \sim N(\beta_1, \tau^2)$.

For right censored data, the information for subject i from cluster j is contained in the pair $(y_{ij},\ \delta_{ij})$ where y_{ij} is the minimum between the event time t_{ij} and the censoring time c_{ij}. δ_{ij} is the censoring indicator, taking the value 1 if the subjects were infected with influenza during the surveillance period, 0 otherwise.

The likelihood of model 6.2.1 is given by:

$$\mathcal{L}(y|\beta_0, \beta_1, b_0, b_1, \phi, \tau) = \prod_{j=1}^{c}\prod_{i=1}^{n_j} \lambda_{ij}(y_{ij})^{\delta_{ij}} \exp\left(-\lambda_{ij}(y_{ij})y_{ij}\right) \tag{6.5.2}$$

and the log-likelihood:

$$\ell(y|\beta_0, \beta_1, b_0, b_1, \phi, \tau) = \sum_{j=1}^{c}\sum_{i=1}^{n_j} \delta_{ij} \log\left(\lambda_{ij}(y_{ij})\right) - \lambda_{ij}(y_{ij})y_{ij} \tag{6.5.3}$$

in which b_0 and b_1 are vectors.

We consider the following standard priors (Smith et al. 1995; Wolfinger 1998):

$\beta_0 \sim N(0, \sigma^2_{\beta_0})$
$\beta_1 \sim N(0, \sigma^2_{\beta_1})$

and hyperpriors:

$$\frac{1}{\phi^2} \sim \Gamma(k_0, l_0)$$
$$\frac{1}{\tau^2} \sim \Gamma(k_1, l_1)$$

6.5.2 Intervals computation

In section 6.4, we saw that in the Bayesian setting, credible, prediction and tolerance intervals were derived from the posterior density of the parameter of interest, the mean vaccine effect, β_1, for credible intervals, a new cluster-specific effect, for prediction intervals and quantiles of the distribution of new cluster-specific effects for tolerance intervals. These posterior densities can either be derived analytically from the joint posterior density $p(\beta_0, \beta_1, \phi, \tau|y)$ or, when no analytical solutions exist, through MCMC simulation methodologies.

The joint posterior density function for the our model is

$$\begin{aligned} p(\beta_0, \beta_1, \phi, \tau|y) \propto & \\ & \prod_{j=1}^{J} \prod_{i=1}^{n_j} [\exp(b_{0,j} + b_{1,j}x_i) \exp(-\exp(b_{0,j} + b_{1,j}x_i)y_i)] \\ & \times \prod_{j=1}^{J} \left[\left[\frac{1}{\phi^2} \exp \frac{-(b_{0,j} - \beta_0)^2}{2\phi^2} \right] \left[\frac{1}{\tau^2} \exp \frac{-(b_{1,j} - \beta_1)^2}{2\tau^2} \right] \right] \\ & \times \left[\exp(\frac{-\beta_0^2}{2\sigma_{\beta_0}^2}) \right] \left[\exp(\frac{-\beta_1^2}{2\sigma_{\beta_1}^2}) \right] \\ & \times \left[\frac{l_0^{k_0}}{\Gamma(k_0)} \phi^{-2(k_0+1)} \exp(\frac{-l_0}{\phi^2}) \right] \left[\frac{l_1^{k_1}}{\Gamma(k_1)} \tau^{-2(k_1+1)} \exp(\frac{-l_1}{\tau^2}) \right] \end{aligned} \tag{6.5.4}$$

where $\sigma_{\beta_0}^2$ is the variance for the parameter β_1, $\sigma_{\beta_0}^2$ the variance for the parameter β_1, k_0 and l_0 are respectively the shape and scale parameters of the inverse gamma distribution for ϕ^2 and k_1 and l_1 the shape and scale parameters of the inverse gamma distribution for τ^2.

The posterior density of b_1 can be obtained as:

$$\begin{aligned} p(b_1|\beta_1, \tau, y) &= p(b_1|\beta_1, \tau)p(\beta_1, \tau|y) \\ &= p(b_1|\beta_1, \tau) \int_{\beta_0, \phi} p(\beta_0, \beta_1, \phi, \tau|y) d\beta_0 d\phi \end{aligned} \tag{6.5.5}$$

In Appendix 6.A, we show that there is no analytical solution to integrate out β_0 out of the joint posterior density for an exponential survival model with no random effect. A fortiori, the same conclusion holds for our model including two random effects. Consequently, instead of relying on complex approximations, MCMC simulations are used to obtain samples from the joint posterior distribution of the parameters.

Credible intervals

To make inference for β_1, we draw a sample of size S of instances of the variable β_1, denoted $\{\beta_1\}^{(S)}$. We then take the empirical $(\frac{\alpha}{2})100\%$ and $(1-\frac{\alpha}{2})100\%$ percentiles of the posterior sample $\{\beta_1\}^{(S)}$ for a $100(1-\alpha)\%$ bilateral interval. To compute an unilateral interval, we take the $100(1-\alpha)\%$ percentile of the posterior sample $\{\beta_1\}^{(S)}$.

Prediction intervals

Since we do not have a closed form for the posterior distribution of β_1, difficulties also arises in the analytical identification of the predictive distribution of b_1. Instead, we apply a simple simulation procedure to draw samples from $p(b_{1,new}|y)$. Following the model, the random effect b_1 is distributed as a Normal with a posteriori location $(\beta_1|y)$ and variance $(\tau^2|y)$. We start by drawing samples of size S of β_1 and τ^2, denoted $\{\beta_1\}^{(S)}$ and $\{\tau\}^{2(S)}$ respectively. We then draw a sample $\{b_1\}^{(S)}$ of new cluster-specific random vaccine effects from $N(\beta_1^{(s)}, \tau^{(s)})$ with $s = 1, ..., S$. The prediction interval is derived by taking the empirical quantiles of this sample.

Tolerance intervals

Derivations of unilateral and bilateral tolerance intervals differ slightly so we present them separately. In both cases, derivation of tolerance interval bound(s) start with the generation of posterior samples for β_1 and τ^2 as for prediction intervals.

In the case of an unilateral tolerance interval, for each pair of $\beta_1^{(s)}$ and $\tau^{2(s)}$, $s = 1, ..., S$, a $(1-p)^{th}$ quantile is derived for the bivariate samples of distributions $N(\beta_1^{(s)}, \tau^{2(s)})$, $s = 1, ..., S$ for a new cluster-specific vaccine effect. This sample can be viewed as a sample from the marginal posterior of q, the $(1-p)^{th}$ quantile of the distribution of a new cluster-specific vaccine effect and inference about the posterior distribution of this quantile can be empirically done based on this sample (Wolfinger 1998).

Construction of a bilateral tolerance interval is more complex. Indeed, we cannot simply compute upper and lower posterior bounds separately and then combine them since the two quantiles do not have a posterior correlation of 1 (Wolfinger 1998).

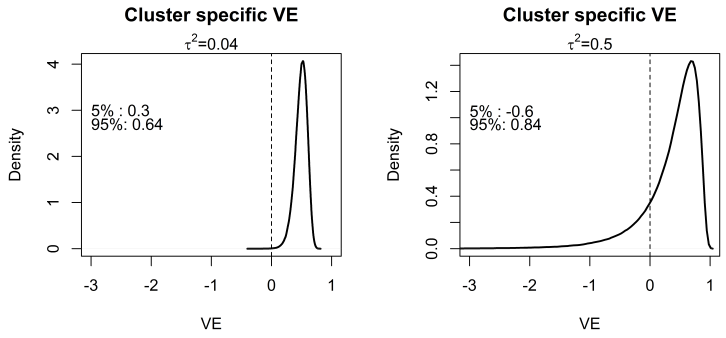

Figure 6.1: Densities of the cluster-specific VE for $\tau^2 = 0.04$ and $\tau^2 = 0.5$ with quantiles 5% and 95%. The dashed line indicates VE=0 (no efficacy).

Instead, we conjointly compute the $(\frac{1-p}{2})^{th}$ and $\frac{p}{2}^{th}$ quantiles from $N(\beta_1^{(s)}, \tau^{2(s)})$, $s = 1, ..., S$. This bivariate sample, denoted $\{(q_l, q_u)\}^{(S)}$ can be viewed as a sample from the marginal posterior of (q_l, q_u), the $(1-p)^{th}$ credible interval for $b_{1,new}$, a new cluster-specific vaccine effect. We then iteratively search for the smallest interval, symmetrical to the mean empirical vaccine effect $\hat{\beta}_1$ and including $100(1-\alpha)\%$ of the $p\%$ posterior intervals derived in the previous step.

6.6 Simulation study

We now run a large simulation study to illustrate and explore the performances of our proposed methodology. We generate CT data with either $J = 5, 10$ or 20 clusters in combination with $n_j = 10, 50$ or 100, yielding to total sample size between 50 and 2000. We simulate 250 datasets for each scenario.

Time-to-event observations, y_{ij} are generated from an exponential survival model with $\beta_0 = 1$ and $\beta_1 = -0.69$ corresponding to $VE = 0.5$. We do not consider any censoring so $\delta_{ij} = 1$ for all observations.

The cluster-specific random variables were generated as follow:

$$b_{0j} \sim N(0, \phi^2) \text{ with } \phi^2 = 0.5$$
$$b_{1j} \sim N(0, \tau^2) \text{ with } \tau^2 = 0.04 \text{ or } 0.5$$

The smaller simulated value for τ^2 was derived from trial Influence 65 while the larger one was chosen to study the impact of large heterogeneity on the analysis results. Figure 6.1 shows the range of cluster-specific VE generated from the two selected distributions.

We implemented our methodology in SAS v9.3 and used the Bayesian procedure PROC MCMC to estimate the model parameters.

SAS code for our model

```
PROC MCMC data=data;
/* initial values */
PARMS beta0 0 beta1 0;
PARMS phi2 1 tau2 1;

/* Random terms definition */
RANDOM b0 ~ normal(mean=beta0, var = phi2)
subject=cluster;
RANDOM b1 ~ normal(mean=beta1, var = tau2)
subject=cluster;

/* Priors */
PRIORS beta0 ~ normal(mean , var);
PRIORS beta1 ~ normal(mean , var);

/* Hyperpriors */
HYPERPRIOR phi2 ~ igamma(shape, scale);
HYPERPRIOR tau2 ~ igamma(shape, scale);

/* Model definition */
linpred = ((b0)+(b1)*vaccine);
llike = event*(linpred) - time*exp(linpred);
model general(llike);
RUN;
```

We used the following vague priors:

$$\beta_0 \sim N(0, 0.00001)$$
$$\beta_1 \sim N(0, 0.00001)$$

$$\frac{1}{\phi^2} \sim \Gamma(0.0001, 0.0001)$$
$$\frac{1}{\tau^2} \sim \Gamma(0.0001, 0.0001)$$

We compute bilateral and unilateral credibility, prediction and 80%-content tolerance intervals. All intervals are computed at the 95% level ($\alpha = 0.05$). More specifically, the tolerance intervals are computed at the 95% level for a content of 80% of future cluster-specific vaccine effects. We generate posterior sample of size $S = 40000$ with a thinning factor of 10. Validation of the of the statistical intervals are computed as the coverage of each interval for 10000 new $\beta_{1,new}$ values drawn from $N(\beta_1, \tau^2)$.

6.6.1 Results

The results of the simulation study are shown in table 6.1. The empirical mean, standard deviation and median for the posterior samples of the parameters of interest are derived. The estimation of the vaccine effect fixed parameter β_1 is close to the true value in all settings. Its precision increases with the number of events by clusters and also the number of clusters. Precision is lower when the inter-cluster heterogeneity is larger. The estimations of τ^2, as reflected by the empirical median seems to be well estimated. Empirical mean of τ^2 tends to be higher than the simulated value for designs with few clusters and subjects, due to the asymmetry of this parameter posterior distribution (Agresti and Finlay 2008).

Figure 6.2 shows the bilateral credibility, prediction and 80%-content tolerance interval limits for each combination of j and n_j for $\tau^2 = 0.04$. Results for $\tau^2 = 0.5$ are similar and not presented. Lower limits of the intervals for VE are presented on the x-axis and upper limits on the y-axis. As expected, p-content tolerance intervals are larger than prediction intervals, which are larger than credibility intervals for the marginal VE. The length of the intervals all decreases with the number of clusters and events. However, the gain of adding clusters seems to be greater than the gain of adding events by cluster. For example, intervals are much smaller with $J = 20$ and $n_j = 20$ ($N = 400$) than for $J = 5$ and $n_j = 100$ ($N = 500$) leading to higher levels of significativity defined as the lower bound of the interval greater than zero.

Table 6.2 shows the mean coverages and proportions of intervals having coverage of at least 80% for 10000 new values of $b_{1,j}$. Figure 6.3 shows the mean coverages (left column) and the proportions of validation samples for which the coverage is at least 80% (right column) by number of clusters, number of subjects by clusters and between clusters vaccine effect heterogeneity levels for the bilateral three types of intervals. Credible intervals are not addressing the question of future cluster-specific VE and as such no specific level of coverage for future vaccine effects is expected. In our work, the prediction intervals address the question of the location of one future vaccine cluster-specific effect. The mean coverages are expected to reach at least the nominal confidence level, here 95%. Finally, p-content α-level tolerance intervals are addressing the question of the location of p$\%$ new cluster-specific vaccine effects. The proportion of validation samples covered in at least 80% of the cases should be at least

95% in our setting. As expected, coverages of tolerance intervals are larger than coverages of prediction intervals, which are larger than coverages of credibility intervals. In some scenarios, coverages of the prediction and tolerance intervals, especially in the unilateral case, are a little smaller than expected, due to Monte-Carlo sampling errors (Lunn et al. 2012). Coverages of the credibility intervals on the mean VE, the classical methodology, decrease rapidly as the number of events and clusters increase. For example, for $J = 10$, $n_j = 100$ and $\tau^2 = 0.04$, a situation similar to that of trial Influence 65, a bilateral credibility intervals only covers 2% of new VE.

τ^2	n by cluster	True value	$J = 5$ Mean(sd)	$J = 5$ Median	$J = 10$ Mean(sd)	$J = 10$ Median	$J = 20$ Mean(sd)	$J = 20$ Median
0.04	**10**	$\beta_0 = 1.00$	1.03 (0.45)	1.03	0.99 (0.26)	0.99	1 (0.18)	1
		$\beta_1 = -0.69$	-0.74 (0.28)	-0.74	-0.7 (0.17)	-0.7	-0.69 (0.11)	-0.69
		$\tau^2 = 0.04$	0.19 (0.66)	0.06	0.09 (0.13)	0.05	0.06 (0.06)	0.03
		$\phi^2 = 0.50$	1.04 (2.68)	0.59	0.64 (0.44)	0.52	0.55 (0.22)	0.51
	20	$\beta_0 = 1.00$	0.99 (0.43)	0.99	1.01 (0.25)	1.01	1 (0.17)	1
		$\beta_1 = -0.69$	-0.7 (0.21)	-0.7	-0.7 (0.13)	-0.7	-0.69 (0.08)	-0.69
		$\tau^2 = 0.04$	0.12 (0.38)	0.04	0.07 (0.09)	0.04	0.04 (0.04)	0.03
		$\phi^2 = 0.50$	1.02 (2.49)	0.59	0.62 (0.42)	0.52	0.55 (0.21)	0.51
	100	$\beta_0 = 1.00$	0.98 (0.42)	0.98	1.02 (0.25)	1.02	1.01 (0.17)	1.01
		$\beta_1 = -0.69$	-0.69 (0.15)	-0.69	-0.7 (0.08)	-0.7	-0.69 (0.05)	-0.69
		$\tau^2 = 0.04$	0.1 (0.29)	0.05	0.05 (0.04)	0.04	0.04 (0.02)	0.04
		$\phi^2 = 0.50$	0.97 (2.74)	0.57	0.66 (0.42)	0.56	0.58 (0.21)	0.53
0.50	**10**	$\beta_0 = 1.00$	1 (0.43)	1	0.99 (0.27)	1	0.99 (0.18)	0.99
		$\beta_1 = -0.69$	-0.66 (0.46)	-0.66	-0.71 (0.28)	-0.71	-0.7 (0.19)	-0.7
		$\tau^2 = 0.50$	1.03 (2.6)	0.53	0.6 (0.51)	0.47	0.57 (0.28)	0.51
		$\phi^2 = 0.50$	0.96 (2.23)	0.53	0.66 (0.47)	0.54	0.54 (0.23)	0.49
	20	$\beta_0 = 1.00$	1.03 (0.43)	1.03	0.99 (0.25)	0.99	0.99 (0.17)	0.99
		$\beta_1 = -0.69$	-0.71 (0.45)	-0.71	-0.69 (0.26)	-0.69	-0.69 (0.18)	-0.69
		$\tau^2 = 0.50$	1.08 (2.53)	0.6	0.61 (0.45)	0.49	0.55 (0.24)	0.5
		$\phi^2 = 0.50$	1.01 (2.16)	0.58	0.64 (0.43)	0.53	0.55 (0.22)	0.5
	100	$\beta_0 = 1.00$	1 (0.42)	1	1.01 (0.25)	1.01	1 (0.17)	1
		$\beta_1 = -0.69$	-0.68 (0.41)	-0.68	-0.69 (0.25)	-0.69	-0.71 (0.17)	-0.71
		$\tau^2 = 0.50$	0.95 (1.94)	0.56	0.66 (0.43)	0.55	0.55 (0.21)	0.51
		$\phi^2 = 0.50$	0.99 (2.17)	0.58	0.62 (0.4)	0.52	0.56 (0.21)	0.52

Table 6.1: Empirical means, standard deviations and medians of the model parameters posterior samples for all the combinations of numbers of clusters and numbers of subjects by clusters for $\tau^2 = 0.04$ and $\tau^2 = 0.5$.

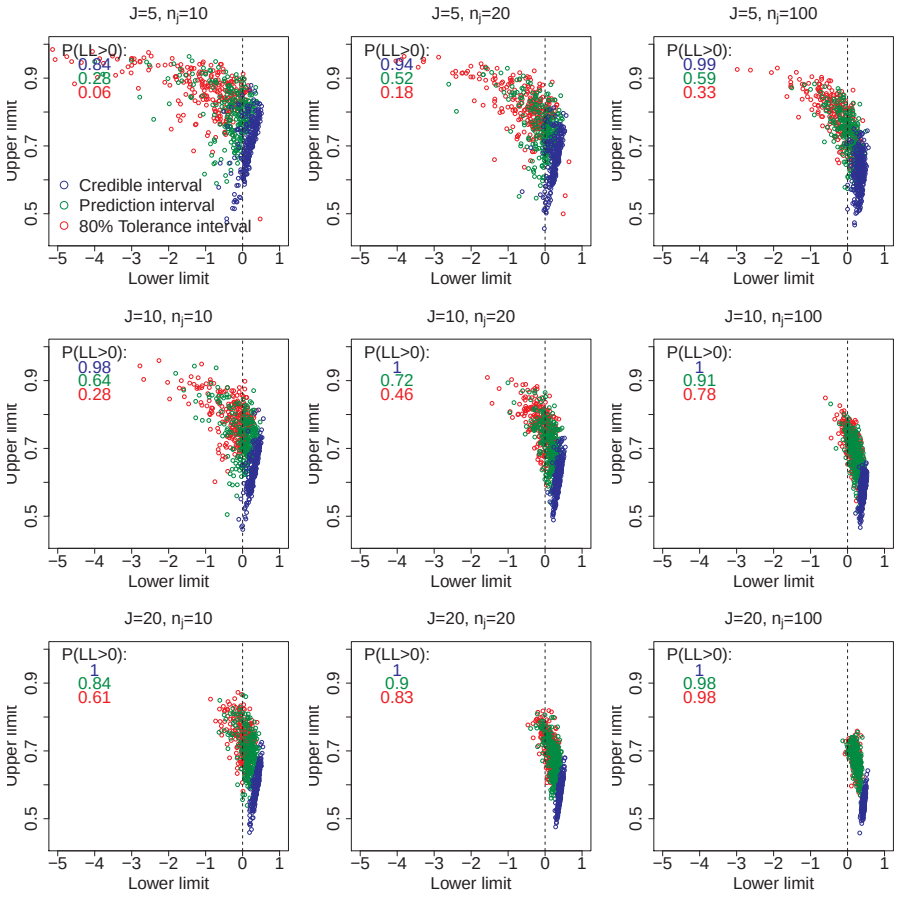

Figure 6.2: Lower and upper bivariate limits and proportions of intervals excluding VE=0 for the three types of intervals for each combination of J and n_j for $\tau^2 = 0.04$.

τ^2	n by cluster	Interval type	$J = 5$ mean cov (≥ 0.8)	$J = 10$ mean cov (≥ 0.8)	$J = 20$ mean cov (≥ 0.8)
0.04	**10**	**CI (bi)**	0.92 (0.87)	0.81 (0.69)	0.67 (0.06)
		CI (uni)	0.9 (0.83)	0.86 (0.75)	0.79 (0.58)
		PI (bi)	0.99 (1)	0.97 (0.99)	0.95 (0.97)
		PI (uni)	0.98 (0.97)	0.96 (0.96)	0.94 (0.94)
		TI (bi)	1 (1)	1 (1)	0.99 (1)
		TI (uni)	0.99 (1)	0.98 (0.99)	0.96 (0.97)
	20	**CI (bi)**	0.85 (0.78)	0.72 (0.21)	0.56 (0)
		CI (uni)	0.88 (0.79)	0.81 (0.63)	0.73 (0.34)
		PI (bi)	0.98 (0.97)	0.96 (0.97)	0.92 (0.89)
		PI (uni)	0.96 (0.96)	0.95 (0.94)	0.92 (0.88)
		TI (bi)	0.99 (0.99)	0.99 (1)	0.97 (0.99)
		TI (uni)	0.98 (0.98)	0.97 (0.98)	0.93 (0.92)
	100	**CI (bi)**	0.75 (0.42)	0.53 (0)	0.4 (0)
		CI (uni)	0.82 (0.64)	0.71 (0.28)	0.67 (0.11)
		PI (bi)	0.96 (0.94)	0.92 (0.88)	0.93 (0.92)
		PI (uni)	0.95 (0.95)	0.92 (0.89)	0.94 (0.97)
		TI (bi)	0.99 (1)	0.96 (0.95)	0.95 (0.97)
		TI (uni)	0.97 (0.97)	0.92 (0.9)	0.91 (0.96)
0.50	**10**	**CI (bi)**	0.68 (0.33)	0.54 (0)	0.4 (0)
		CI (uni)	0.8 (0.59)	0.71 (0.29)	0.67 (0.1)
		PI (bi)	0.9 (0.76)	0.91 (0.86)	0.94 (0.94)
		PI (uni)	0.92 (0.85)	0.91 (0.88)	0.95 (0.99)
		TI (bi)	0.96 (0.96)	0.95 (0.92)	0.95 (0.97)
		TI (uni)	0.94 (0.91)	0.92 (0.9)	0.92 (0.96)
	20	**CI (bi)**	0.69 (0.37)	0.51 (0.01)	0.37 (0)
		CI (uni)	0.78 (0.56)	0.71 (0.26)	0.65 (0.04)
		PI (bi)	0.91 (0.82)	0.93 (0.91)	0.94 (0.97)
		PI (uni)	0.92 (0.86)	0.94 (0.93)	0.94 (0.98)
		TI (bi)	0.96 (0.91)	0.96 (0.94)	0.94 (0.97)
		TI (uni)	0.94 (0.9)	0.93 (0.94)	0.9 (0.94)
	100	**CI (bi)**	0.68 (0.32)	0.5 (0)	0.35 (0)
		CI (uni)	0.79 (0.56)	0.71 (0.2)	0.64 (0)
		PI (bi)	0.94 (0.9)	0.95 (0.96)	0.94 (0.99)
		PI (uni)	0.95 (0.93)	0.95 (0.98)	0.95 (1)
		TI (bi)	0.98 (0.97)	0.97 (0.98)	0.94 (0.99)
		TI (uni)	0.96 (0.94)	0.93 (0.96)	0.9 (0.96)

Table 6.2: Mean coverages and proportions of intervals covering at least 80% of 10000 new cluster-specific VE.

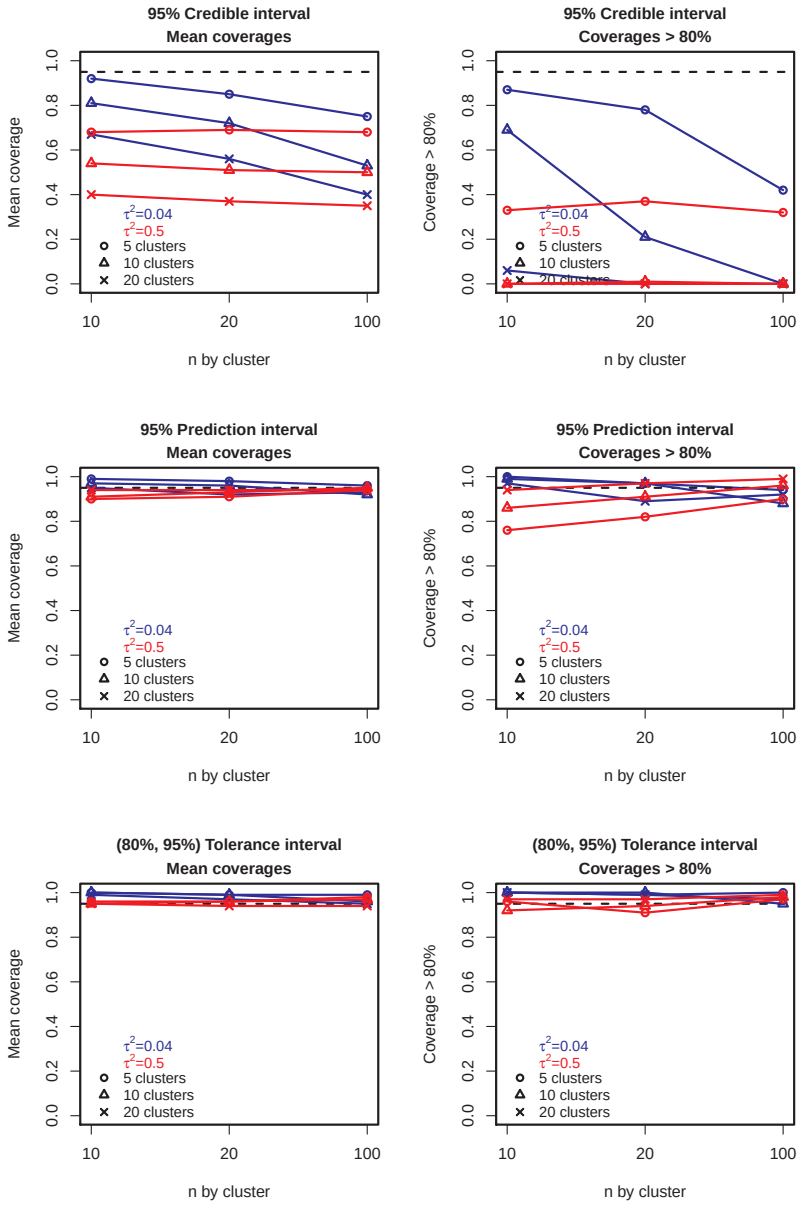

Figure 6.3: Mean coverages and proportions of intervals covering at least 80% of 10000 new cluster-specific VE by bilateral type of interval, for each combination of j and n_j for $\tau^2 = 0.04$ and $\tau^2 = 0.5$.

6.7 Reanalysis of trial Influence 65

In this section, we re-analyse the trial Influence 65 according to the proposed methodology. The conclusion of this trial was that there was no evidence of increased VE of the adjuvanted vaccine over the standard vaccine when considering all influenza strains (see Chapter 4 for more details), resulting in little heterogeneity across country-specific VE. However, the efficacy results for the flu strain A/H3N2 seem more encouraging. For this reason, we decided to only consider the clinical influenza cases caused by strain H3N2 for our reanalysis. VE heterogeneity across countries here could be caused by virus mutations or varying levels of vaccine coverage for example.

We performed two analyses considering two levels of clustering: countries ($J = 15$) and, for the purpose of illustration, also considered a somewhat more artificial clustering into geographical regions ($J = 4$). We analyse the data with the model given in equation 6.5.1 and we use standard vague priors as described in section 6.5.1. We derive unilateral and bilateral $100(1-\alpha)\%$ level credible intervals, PI and 80% content TI.

Model estimates for both clustering scenarios are presented in table 6.3. For the same total number of events, estimations are much more precise in the by-country setting. Figure 6.4 and 6.5 show the cluster-specific VE (with 95% credible interval) as well as a 95% bilateral credible interval for the marginal VE, 95% bilateral prediction interval for a new cluster-specific VE and a 80% content 95% confidence level tolerance interval. The additional precision on both the estimations of β_1 and τ^2 for the analysis considering country clustering translates into much smaller statistical intervals, especially for the prediction interval and the tolerance interval which include imprecision information also on τ^2.

	Geographical region		**Country**	
Parameter	**Mean (sd)**	**Median**	**Mean (sd)**	**Median**
$\boldsymbol{\beta_0}$	-10.48 (1.21)	-10.41	-9.9 (0.3)	-9.88
$\boldsymbol{\beta_1}$	-0.39 (0.55)	-0.32	-0.26 (0.11)	-0.25
$\boldsymbol{\tau^2}$	0.72 (6.74)	0.05	0.04 (0.06)	0.03
$\boldsymbol{\phi^2}$	5.25 (13.18)	2.29	1.15 (0.58)	1.04

Table 6.3: Empirical means, standard deviations and medians of the model parameters posterior samples for the reanalysis of trial Influence 65 for the geographical regions and the countries clustering levels.

Cluster	Marginal VE	Credible interval	Prediction interval	80 % Tolerance interval
Region	0.32	-0.09, 0.71	-0.74, 0.88	-3.86, 0.91
Country	0.23	0.05, 0.3	-0.21, 0.54	-0.38, 0.57

Table 6.4: 95% credible intervals for the marginal VE, prediction interval for a new cluster-specific VE and 80% content tolerance interval for new cluster-specific VE for the reanalysis of trial Influence 65 and by clustering level (Geographical regions or countries).

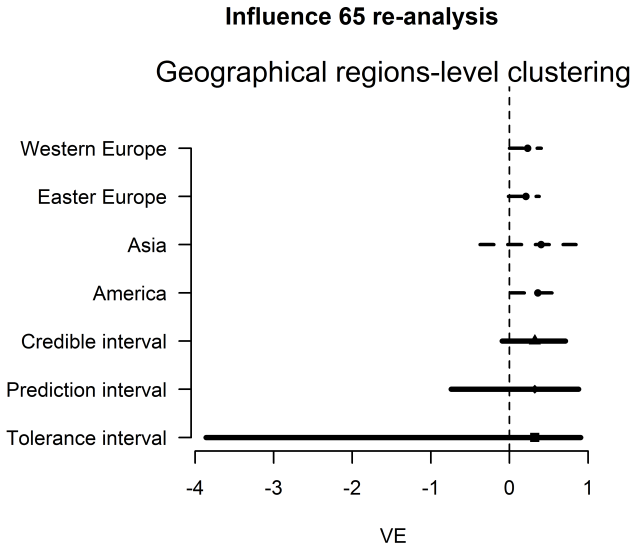

Figure 6.4: VE by geographical region (with 95% credible intervals), marginal VE (with 95% credible interval), 95% prediction interval for a new cluster-specific VE and 80%-content 95% tolerance interval for new cluster-specific VE. The dashed line represents VE=0.

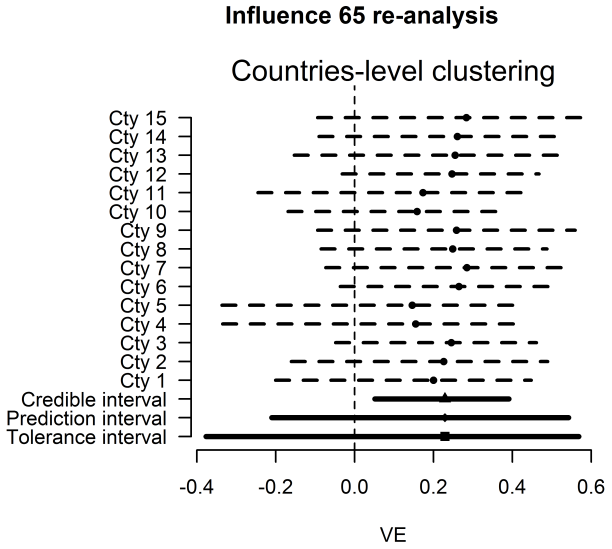

Figure 6.5: VE by country (with 95% credible intervals), marginal VE (with 95% credible interval), 95% prediction interval for a new cluster-specific VE and 80%-content 95% tolerance interval for new cluster-specific VE. The dashed line represents VE=0.

The reanalysis of the trial Influence 65 provides us with information about the VE heterogeneity and about the VE of the adjuvanted vaccine compared to the standard of care in the future. While the credible interval for the marginal VE gives information about the mean VE observed in the trial, the 95% level prediction interval constructed from the country clustering model indicates that if, for example, we were to run a new trial within one new cluster, the new VE would be between -0.21 and 0.54, with confidence level 95%. The 80%-content tolerance 95% level tolerance intervals indicates that if we were to use the product in 10 new clusters, 8 of the cluster-specific VE would be between -0.38 and 0.57, with confidence level 95%.

While predictive intervals are larger than credible intervals for the marginal mean, as shown by the comparison of the coverages for new cluster-specific VE by the three types of intervals in the previous section,they also provide a much more complete answer to the question of VE in new clusters, i.e. regions and season and thus directly address the objectives of a phase III seasonal influenza VE trial.

6.8 Discussion

Classically, seasonal influenza CT data are analysed with a model assuming constant vaccine effect between clusters (countries and flu seasons) and decisions on the efficacy of the tested vaccine are based on confidence intervals for this overall efficacy. In this chapter, we argue that this methodology does not address the efficacy of the tested vaccine in the future, i.e. in new clusters.

We propose an analysis model accounting for VE heterogeneity for seasonal influenza. In combination with our model, we suggest the use of predictive intervals instead of confidence intervals for a mean VE. Indeed, the coverage of a CI for new cluster-specific VE is very low for design and heterogeneity levels expected in phase III trials. This alone could explain why the same product could not be proved efficacious in one trial and then show significant efficacy in a new trial (Beran et al. 2009b,a).

In this chapter, we have explored the use of prediction and p-content tolerance intervals. While the first one gives information about the probable location of one new cluster-specific VE, the second gives information about the location of simultaneous p% of new cluster-specific VE. Because the seasonal influenza vaccine is expected to be used in numerous new clusters, p-content tolerance intervals are more adapted.

Predictive intervals are larger than confidence intervals. Their lengths depends upon the quantity of information about the mean VE, the heterogeneity between clusters, and the information about this heterogeneity, i.e. the number of events and the number of clusters. We showed that increasing the number of clusters provided more information (thus more precision in the estimation of the model parameters) than increasing the size of the clusters. Our recommendation is to run new trials in multiple

countries and over several seasons to collect sufficient information to have a precise estimation of the inter-cluster heterogeneity.

While our methodology provides a better answer to question of future VE, for a similar success criteria definition, the disadvantage is an increase of the trial size. Determination of the design and size of trials to be analysed with our methodology requires further research. From Equation 6.5.4 and based on our simulations results, it appears that both the number of events and the number of clusters have to be taken into account in this process. Ideas from the work realized in the context of variance component linear models, such as Hugo (2012) and Lebrun (2012), may be helpful in this regards. Because predictive intervals address different questions than the classically used confidence intervals, success criteria for a trial may be modified. This particular point should be discussed with the pharmaceutical authorities.

Our proposal also bring some interesting prospects from an exploratory point of view. For example, by studying the heterogeneity of VE between countries, the impacts of factors such as vaccination policies, people influenza vaccine history or the effects of non influenza immunizations on VE could be explored.

Before application to a new phase III trial however, more research is necessary. While our methodology can be applied in various contexts and in conjunction with different analysis models, we made a number of simplifications for the validation part of our work. First, we have considered the special case of no censoring and balanced repartition of events across clusters. This assumption is highly unlikely in the context of seasonal influenza phase III trials and should be relaxed in future validation research. However, the results should not be impacted. Second, our analysis model is a completely parametrical survival regression model which does not reflect some of the specificities of seasonal influenza. Our methodology should be validated using more complex models such as semi-parametrical Cox regression mixed models (Çetinyürek Yavuz and Lambert 2011; Corbière et al. 2009) or piecewise-constant hazard frailty models (Rondeau et al. 2006). These kind of models present a new complexity since they add source of randomness at the baseline hazard level, with also a cluster-specific attribute. Nevertheless, because the AR of seasonal influenza tends to be small, for sample sizes similar to what is currently done, the data tend to be highly censored and fitting of complex model tends to be challenging.

In our work, we considered different clustering levels, as illustrated in the reanalysis of trial Influence 65: geographical region and country. We have shown that, for the same number of events, there was a statistical gain in having more clusters. However, we feel that the definition of the clustering level should not be a statistical decision but that it should be defined a priori to reflect what is clinically relevant. Influenza geographical regions could for example be used. In our work, we have considered geographical (country or geographical region) at the same level as season. This as-

sumption should be explored in the future and future trials ran over several influenza seasons.

Finally, we have used non informative priors. It could be reasonable to take advantage of the Bayesian setting to include a priori knowledge about VE heterogeneity into the analysis of phase III clinical trials. Including prior knowledge in the model could lead to shorter intervals and potentially smaller sample sizes to reach a similar power level as compared to the case of non-informative priors. Unfortunately, there is little information available so far, especially about the season effect since trials tend to be run over a single season. Whether informative priors could be retrieved from immunogenicity trials, which are run every season, could be explored.

Summary of Chapter 6

- Because of strains heterogeneity, the classical decision paradigm for phase III trials provides an incomplete picture of vaccine efficacy in the future.
- Our new methodology allows the characterisation of vaccine efficacy heterogeneity across clusters by including a random vaccine-by-cluster interaction effect.
- The use of predictive (prediction or tolerance) intervals instead of confidence intervals to answer the question of future VE provide insight on the range of future VE across seasons and geographical regions and is much more relevant in the case of heterogeneous diseases.

6.A Analytical derivations of the marginal posterior density for β_1

We consider a completely parametrical model in which the survival times y follow an exponential distribution with parameter λ_0. This model implies that the baseline hazard λ_0 is constant over time. We have:

$$h(t|X = x) = \lambda_0 \exp(\beta_1 x) \ \forall\, t \in [0, +\infty[\tag{6.A.1}$$

with $t \sim \exp(\lambda_0)$ and $h_0(t) = \lambda_0$.

We re-parametrize this model as

$$h(t|X = x) = \exp(\beta_0 + \beta_1 x) \ \forall\, t \in [0, +\infty[\tag{6.A.2}$$

where $\lambda_0 = \exp(\beta_0)$.

This model has the following survival function:

$$S(t|X = x) = \exp\left[-\exp(\beta_0 + \beta_1 x)t\right] \tag{6.A.3}$$

The likelihood function is

$$\mathcal{L}(y_i; \theta) = \prod_{i=1}^{n} \exp(\beta_0 + \beta_1 x_i) \exp(-\exp(\beta_0 + \beta_1 x_i) y_i) \tag{6.A.4}$$

We let the two unknown parameters β_0 and β_1 take on the following normal prior distributions:

$$\begin{aligned} \beta_0 &\sim N(0, \sigma^2_{\beta_0}) \\ \beta_1 &\sim N(0, \sigma^2_{\beta_1}) \end{aligned} \tag{6.A.5}$$

Bayesian inference is based on the posterior distribution which is obtained by multiplying the likelihood function with the prior density functions as given below:

$$\begin{aligned}
&\pi(\beta_0, \beta_1|y) \\
&\propto \prod_{i=1}^{n} [\exp(\beta_0 + \beta_1 x) \exp(-\exp(\beta_0 + \beta_1 x_i) y_i)] \exp(\frac{-\beta_0^2}{2\sigma_{\beta_0}^2}) \exp(\frac{-\beta_1^2}{2\sigma_{\beta_1}^2}) \\
&\propto \exp\left[n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i\right] \exp\left[-\exp(\beta_0) \sum_{i=1}^{n} \exp(\beta_1 x_i) y_i\right] \exp(\frac{-\beta_0^2}{2\sigma_{\beta_0}^2}) \exp(\frac{-\beta_1^2}{2\sigma_{\beta_1}^2})
\end{aligned} \tag{6.A.6}$$

We seek to determine the posterior marginal distribution for β_1. For that, we integrate out the nuisance parameter β_0

$$\begin{aligned}
\int_{-\infty}^{+\infty} \pi(\beta_0, \beta_1|y) d\beta_0 &= \exp\left[\beta_1 \sum_{i=1}^{n} x_i - \frac{\beta_1^2}{2\sigma_{\beta_1}^2}\right] \\
&\times \int_{-\infty}^{+\infty} \left\{ \exp\left[n\beta_0 - \exp(\beta_0) \sum_{i=1}^{n} (\exp(\beta_1 x_i) y_i) - \frac{\beta_0^2}{2\sigma_{\beta_0}^2}\right] \right\} d\beta_0
\end{aligned} \tag{6.A.7}$$

The following integral has to be solved:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \pi(\beta_0, \beta_1|y) d\beta_0 \\
&= \int_{-\infty}^{+\infty} \left\{ \exp\left[n\beta_0 - \frac{\beta_0^2}{2\sigma_{\beta_0}^2} - \exp(\beta_0) \sum_{i=1}^{n} (\exp(\beta_1 x_i) y_i)\right] \right\} d\beta_0 \\
&= \int_{-\infty}^{+\infty} \left\{ \exp\left[-\frac{1}{2\sigma_{\beta_0}^2} \left(\beta_0^2 - 2\sigma_{\beta_0}^2 \beta_0 \left(n - \frac{\exp(\beta_0)}{\beta_0} \sum_{i=1}^{n} (\exp(\beta_1 x_i) y_i)\right)\right)\right] \right\} d\beta_0
\end{aligned} \tag{6.A.8}$$

Unfortunately, there is no analytical solution for this equation and the posterior marginal distribution of β_1 has to be derived through numerical methodologies.

Chapter 7

Conclusion

*

Concluding discussion

Influenza is an infectious disease which seasonal epidemics cause an estimated 250000 to 500000 deaths yearly (WHO 2012). Annual vaccination is therefore recommended, especially in the high risk populations. Unfortunately, immunosenescence and mismatching between vaccine and circulating strains are linked to decrease VE. To overcome these issues, pharmaceutical companies are developing novel seasonal influenza vaccines that offer better protection. The start point of this thesis is the observation that many seasonal influenza phase III large clinical trials failed to show significant VE (Dewé et al. 2013). Moreover, it also occurred that the same vaccine could not be proved efficacious in a first trial but then showed significant efficacy, according to the classical estimation and design methodology, in a later trial (Beran et al. 2009a,b).

Our project had two main objectives: the first objective was to understand the particularities of seasonal influenza and the context in which VE trials took place. The second objective was to use this information to improve the design of future efficacy trials, to propose a new way of thinking the design of future efficacy trials and to propose more appropriate statistical tools for the analyses of such trials.

We answered these objectives from different axes. In Chapter 2, we attempted to have a better understanding of seasonal influenza and vaccination against this infectious disease. For that, we reviewed over 100 trials. We then identified several potential issues that we explored in the subsequent chapters.

The relevance of post-vaccination HI titres as a COP for the development of seasonal influenza vaccine was discussed in Chapter 3. An exploratory analysis of a pooling of four trials showed that post-vaccination HI titres were significantly associated with the risk of developing clinical influenza. However, we also found out that an absolute threshold for this COP was not relevant since risks of infection with seasonal influenza

depended also on uncontrollable and unobserved factors such as exposure to the virus. In the model developed in Chapter 3 we included season strength as an indicator of the level of exposure. In Chapter 4 we presented a simulation framework that accounts for varying exposure levels between countries, through historical data and contacts behaviours. The analysis methodology we propose in Chapter 6 accounts for exposure heterogeneity between countries.

Combining simulations and analytical results, we then explored the limit of the current methodologies for designing phase III efficacy trials (Chapter 4) and analysing the collected data (Chapter 5). To reach this goal, we had to develop a simulation framework for the generation of phase III clinical trials time-to-infection data. Our methodology is presented in Chapter 4 and accounts for subjects heterogeneity in fragility and in contact rates, two possible mechanisms of vaccine protection and strain heterogeneity in intensity and time course between countries.

In Chapter 4 we showed that small departure from trial hypotheses, especially at the strains-related level, was associated with large decreases in power to show significant VE. This results gives us a better understanding of the potential causes for the lack of significant efficacy observed in recent trials and also in the discordances observed in other cases. In Chapter 5, we investigated the robustness of the classical regression model used to analyse seasonal influenza VE trial data. We showed that in this specific context, because the AR are particularly small, simple models not accounting for sources of heterogeneity still gave good estimates and precision for VE. We also showed that more complex models did not always give better results and, in some cases, were even not easily applicable due to the characteristics of influenza. So, despite the complexity of the influenza context, analysing VE trial with simple models seems to be adequate.

Finally, in Chapter 6, we put in question the actual objectives and decision-making process of phase III trials and proposed a totally new way of approaching the design and the analysis of those trials. We challenged the use of confidence intervals for mean VE and we suggested the use of analysis models allowing heterogeneity in VE across countries and seasons and the derivation of predictive intervals instead of confidence intervals.

Practical perspectives

While our work is directly related to the conduct of phase III VE trials, we feel that the issue is more global and, in hindsight, what we have studied here can be integrated from the very beginning of product development.

First, to develop a pharmaceutical product, the targeted pathology should be characterized and the expected impact of the product well defined (Burman et al. 2005; Burman

and Wiklund 2011). This work and the full development of the product should involve a team of physicians, pharmacologists, laboratory researchers, epidemiologists and statisticians among others. In Chapter 2 we tried to define the dynamics of seasonal influenza and influenza vaccine. Knowledge about any pathology is not static but instead requires continuous updating to reflect the latest information and evidence.

Second, the intermediate and final objectives to achieve should be clearly defined. As the product development advances and depending on the state of the market for the product, objectives should be re-evaluated. Uncertainties and risks about the development of the product should be identified as soon as possible. In Chapter 2, we identified several sources of uncertainties and potential issues with the development of a seasonal influenza vaccine. Ways to reduce these uncertainties should be integrated into the product development plan. For example, in the case of vaccine against seasonal influenza, identification and specification of COP should be kept in mind at all development phases. In Chapter 3, we pooled data from four efficacy trials that used different seasonal influenza vaccines to gain information about post vaccination HI titres as a COP. This information will then be re-injected and updated to develop new vaccines against seasonal influenza. For example, we confirmed that a four-fold pre-post vaccination titre increase was associated with a two-fold decrease in the risk of being infected with seasonal influenza for the virus strain A/H3N2. This link should be confirmed in novel trials and studied for other virus strains. While collecting post-vaccination blood sample in all phase III trials participants may be costly and/or not feasible, substantial amount of information may be achieved by combining data from different pharmaceutical companies.

Uncertainties should also be taken into account in designing the stages of a CT. O'Hagan et al. (2005) point out that powering of a CT is conventionally done conditionally on an expected fixed treatment effect. Because there is no guarantee that the true underlying effect will be equal to the assumed value, they state that power does not quantify the probability of a successful trial. Instead, they recommend the use of an assurance strategy. They define assurance as "the unconditional probability that the trial will yield a positive outcome". The simulation methodology presented in Chapter 4 can be seen as an assurance tool. Here, the uncertainty about VE against seasonal influenza is explored from a mechanistic, epidemiology-inspired model perspective. By envisaging different scenarios, we can test the robustness of the chosen strategy to the identified risks. This additional step may be time-consuming but the benefit would be better probabilities of success of future trials and therefore may eventually decrease the time-to-licensure.

When designing a life development plan and its trials, an assurance strategy requires the integration of many factors and extensive information about both the disease and the product. At the analysis stage however, we recommend a very different strategy, adopting the simplicity principle: the objective is not to fit the perfect model, matching

all the characteristics of the data generation process, but to use a statistical model that gives reliable estimations of the quantity of interest, in this case VE. In Chapter 5 we showed that analysis regression models omitting sources of heterogeneity gave good estimates of VE in the specific context of infectious diseases characterized by low AR.

Finally, in order to answer the questions of CT, we firmly believe that uncertainties should be integrated in the decision-making stage. For a phase III trial, the main objective should be to show the efficacy of the product in new subjects. In Chapter 6, we recommended the use of predictive intervals instead of confidence intervals for this purpose. We showed that confidence intervals gave an incomplete answer to this objective and that statistically significant mean VE in one trial did not ensure VE in new countries and seasons. On the opposite, predictive intervals and more particularly tolerance intervals provide information about VE in the future. While this methodology is associated with larger and longer clinical trials, it also provides better insight on the range of future vaccine efficacies and, as such, provides the patients and physicians with a more complete knowledge of the product. For phase III trials, the criteria of success based on predictive intervals should be re-discussed with the health authorities as they address a different question than the confidence intervals. VE heterogeneity across countries and seasons should also be assessed after a novel vaccine has been released to the market and should therefore be included as an objective of phase IV trials.

Further development

The work started in this thesis was ambitious. While we hope to have opened some doors towards new VE CT methodology, further work is necessary.

Here, we have only considered the traditional two-arms parallel design. However, innovative protocols and designs could be explored. Indeed, in large vaccine efficacy trials, the amount of information collected on each subject is quite limited. The type of vaccine administered, prior vaccination history, location (country, center), gender, age and pre-existing conditions are the usually collected variables. As a result, very little is known about the subject susceptibility to the vaccine-targeted infection, resulting in important non-explained heterogeneity when modelling the infection risks. One important unknown quantity is the level of exposure to the infection undergone by a subject. Ideally, one would like to estimate vaccine efficacy conditionally on virus exposure, such as in a challenge studies. To reduce uncertainty about the level of exposure, covariates that could be indicators of contacts (overall) or contacts with infected people should be collected: jobs, family sizes, use of collective transportations are examples of such variables. Also, the typical number of daily contacts should be collected, at least in a subset subjects who would be asked to fill in a contact diary

(Mossong et al. 2008) for a short period. Finally, clinical trials should include information on small clusters of subjects, such as households, school-aged children attending the same classes, colleagues sharing the same office at work and so on. From the moment member within a cluster becomes infected, all other cluster members can be considered as exposed to the virus and would be monitored closely. Vaccine efficacy could thus be estimated conditionally on exposure. Such trials, including household contact information, have been conducted in order to study antiviral drug efficacy (Halloran et al. 2007).

To control for subject exogenous characteristics, CT should be run over several seasons through a cross-over design. Subjects would be vaccinated during one season and not the other, and as such would be their own control. This design has been used, for example, in the context of other vaccines studies (Park et al. 2004). However, this design would be impossible to apply in a population in whom vaccination is recommended as the standard of care by the health authorities, such as the elderly in the case of influenza and pneumococcal vaccines.. Also, the effect of vaccine history is controversial in the seasonal influenza context (Hoskins et al. 1979; Beyer et al. 1999) and pharmaceutical companies usually prefer to test their vaccines in a naive population (Keitel et al. 1997), i.e. subjects who did not receive the vaccine previously. Brown and Lilford (2006) propose a design called "stepped wedge trial" in which a treatment is rolled-out sequentially to the trial participants over a number of time periods. In this design, subjects who have received the experimental treatment do not go back to the comparator. In all cases, careful care should be taken to control for the differences in levels of exposure between seasons.

In our work, we have considered laboratory-confirmed influenza infection as the only endpoint. However, this endpoint might be ameliorated. For example, the severity of the disease should be considered. Indeed, a significant reduction of the severity of the disease through vaccination might be an appealing vaccine benefit since it would mean lower burden for the sick person but might also lead to reduction of the contagiousness and thus indirect protection of the community. Samples obtained from trial participants are only sent for analysis when subjects experienced clinical manifestations of influenza. However, it is likely that some subjects would be infected without experiencing any symptoms. Nichol (2006) proposes a discussion about the impact of heterogeneity in case-definitions for influenza in the interpretation of trial results.

Finally, while we have suggested to extend phase III seasonal influenza VE trials over several seasons, we have not explored the monitoring strategies for interim analyses and decision-making methodologies that could be applied to this design. Among other possibilities, stochastic curtailment (Jennison and Turnbull 2010) seems an appealing solution. In this methodology, conditional power is used as a basis for stopping the trial. The trial is terminated, for efficacy or lack of efficacy, as soon as a particular decision is highly likely given the current data (Davis and Hardy 1994). Lin et al.

(1999) provide a general theory for applying stochastic curtailment to survival data. Harvey and Kagerer (2005) and Dimitrienko and Wang (2006) provide solutions for the monitoring of clinical trials through Bayesian statistics. However, to gain information about season heterogeneity, we do not recommend to stop trials too early. The situation where there appears to be a lack of vaccine efficacy should also be discussed extensively as lack of efficacy in some clusters might be the results of VE heterogeneity.

Bibliography

Agresti, A. (1996), *An introduction to categorical data analysis*, vol. 135, Wiley New York.

Agresti, A. and Finlay, B. (2008), *Statistical Methods for the Social Sciences (4th Edition)*, Pearson.

Antia, R., Regoes, R. R., Koella, J. C., and Bergstrom, C. T. (2003), "The role of evolution in the emergence of infectious diseases," *Nature*, 426, 658–661.

Atkinson, W., Wolfe, S., and Hamborsky, J. (2011), *Epidemiology and prevention of vaccine-preventable diseases*, Public Health Foundation.

Barker, J., Stevens, D., and Bloomfield, S. (2001), "Spread and prevention of some common viral infections in community facilities and domestic homes," *Journal of Applied Microbiology*, 91, 7–21.

Belongia, E. A., Sundaram, M. E., McClure, D. L., Meece, J. K., Ferdinands, J., and VanWormer, J. J. (2014), "Waning vaccine protection against influenza A (H3N2) illness in children and older adults during a single season," *Vaccine*.

Belshe, R. B., Gruber, W. C., Mendelman, P. M., Cho, I., Reisinger, K., Block, S. L., Wittes, J., Iacuzio, D., Piedra, P., Treanor, J., et al. (2000), "Efficacy of vaccination with live attenuated, cold-adapted, trivalent, intranasal influenza virus vaccine against a variant (A/Sydney) not contained in the vaccine," *The Journal of pediatrics*, 136, 168–175.

Belshe, R. B., Mendelman, P. M., Treanor, J., King, J., Gruber, W. C., Piedra, P., Bernstein, D. I., Hayden, F. G., Kotloff, K., Zangwill, K., et al. (1998), "The efficacy of live attenuated, cold-adapted, trivalent, intranasal influenzavirus vaccine in children," *New England Journal of Medicine*, 338, 1405–1412.

Bender, R., Augustin, T., and Blettner, M. (2005), "Generating survival times to simulate Cox proportional hazards models," *Statistics in medicine*, 24, 1713–1723.

Benoit, A., Beran, J., Devaster, J.-M., Esen, M., Launay, O., Leroux-Roels, G., McElhaney, J. E., Oostvogels, L., Ruiz-Palacios, G., van Essen, G. A., Gaglani, M., Jackson, L., Vesikari, T., Legrand, C., Tibaldi, F., Innis, B., and Dewé, W. (2014a), "Hemagglutination inhibition antibody titers as a correlate of protection against seasonal A/H3N2 influenza disease," *Submitted*.

Benoit, A., Legrand, C., and Dewé, W. (2014b), "Performances of regression models in estimating communicable infectious diseases vaccine efficacy," *Submitted*.

— (Conditionnaly accepted), "Influenza vaccine efficacy trials: a simulation approach to understand failures from the past," *Pharmaceutical Statistics*.

Beran, J., Vesikari, T., Wertzova, V., Karvonen, A., Honegr, K., Lindblad, N., Van Belle, P., Peeters, M., Innis, B. L., and Devaster, J.-M. (2009a), "Efficacy of inactivated split-virus influenza vaccine against culture-confirmed influenza in healthy adults: a prospective, randomized, placebo-controlled trial," *Journal of Infectious Diseases*, 200, 1861–1869.

Beran, J., Wertzova, V., Honegr, K., Kaliskova, E., Havlickova, M., Havlik, J., Jirincova, H., Van Belle, P., Jain, V., Innis, B., et al. (2009b), "Challenge of conducting a placebo-controlled randomized efficacy study for influenza vaccine in a season with low attack rate and a mismatched vaccine B strain: a concrete example," *BMC infectious diseases*, 9, 2.

Beran, R. (1981), "Nonparametric regression with randomly censored survival data," Tech. rep., Technical Report, Univ. California, Berkeley.

Berkson, J. and Gage, R. P. (1952), "Survival curve for cancer patients following treatment," *Journal of the American Statistical Association*, 47, 501–515.

Beyer, W. E., de Bruijn, I. A., Palache, A. M., Westendorp, R. G., and Osterhaus, A. D. (1999), "Protection against influenza after annually repeated vaccination: a meta-analysis of serologic and field studies," *Archives of internal medicine*, 159, 182–188.

Boag, J. W. (1949), "Maximum likelihood estimates of the proportion of patients cured by cancer therapy," *Journal of the Royal Statistical Society. Series B (Methodological)*, 11, 15–53.

Bretagnolle, J. and Huber-Carol, C. (1988), "Effects of omitting covariates in Cox's model for survival data," *Scandinavian Journal of Statistics*, 125–138.

Bridges, C. B., Thompson, W. W., Meltzer, M. I., Reeve, G. R., Talamonti, W. J., Cox, N. J., Lilac, H. A., Hall, H., Klimov, A., and Fukuda, K. (2000), "Effectiveness and cost-benefit of influenza vaccination of healthy working adults: a randomized controlled trial," *Jama*, 284, 1655–1663.

Brown, C. A. and Lilford, R. J. (2006), "The stepped wedge trial design: a systematic review," *BMC medical research methodology*, 6, 54.

Burman, C.-F., Hamren, B., and Olsson, P. (2005), "Modelling and simulation to improve decision-making in clinical development," *Pharmaceutical Statistics*, 4, 47–58.

Burman, C.-F. and Wiklund, S. J. (2011), "Modelling and simulation in the pharmaceutical industry : some reflections," *Pharmaceutical statistics*, 10, 508–516.

Burnham, K. P. and Anderson, D. R. (2004), "Multimodel inference understanding AIC and BIC in model selection," *Sociological methods & research*, 33, 261–304.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006), "The design of simulation studies in medical statistics," *Statistics in medicine*, 25, 4279–4292.

Callas, P. W., Pastides, H., and Hosmer, D. W. (1998), "Empirical comparisons of proportional hazards, poisson, and logistic regression modeling of occupational cohort data," *American journal of industrial medicine*, 33, 33–47.

Carlin, B. P. and Louis, T. A. (2000), *Bayes and empirical Bayes methods for data analysis*, CRC Press.

Carrat, F. and Flahault, A. (2007), "Influenza vaccine: the challenge of antigenic drift," *Vaccine*, 25, 6852–6862.

CDC (2011), "Prevention and control of influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP)," *MMWR. Morbidity and mortality weekly report*, 60, 1128.

Çetinyürek Yavuz, A. and Lambert, P. (2011), "Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized B-splines," *Statistics in medicine*, 30, 75–90.

Chan, I. S., Wang, W. W., and Heyse, J. (2003), "Vaccine clinical trials," *Encyclopedia of Biopharmaceutical Statistics*, 2, 1005–1022.

Chao, D. L., Halloran, M. E., Obenchain, V. J., and Longini Jr, I. M. (2010), "FluTE, a publicly available stochastic influenza epidemic simulation model," *PLoS computational biology*, 6, e1000656.

Clements, M., Betts, R., and Murphy, B. (1984), "Advantage of live attenuated cold-adapted influenza A virus over inactivated vaccine for A/Washington/80 (H3N2) wild-type virus infection," *The Lancet*, 323, 705–708.

Coburn, B. J., Wagner, B. G., and Blower, S. (2009), "Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)," *BMC medicine*, 7, 30.

Congdon, P. (2007), *Bayesian statistical modelling*, vol. 704, John Wiley & Sons.

Cooray, K. (2006), "Generalization of the Weibull distribution: the odd Weibull family," *Statistical Modelling*, 6, 265–277.

Corbière, F., Commenges, D., Taylor, J. M., and Joly, P. (2009), "A penalized likelihood approach for mixture cure models," *Statistics in medicine*, 28, 510–524.

Corbière, F. and Joly, P. (2007), "A SAS macro for parametric and semiparametric mixture cure models," *Computer methods and programs in biomedicine*, 85, 173–180.

Coudeville, L., Bailleux, F., Riche, B., Megas, F., Andre, P., and Ecochard, R. (2010), "Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model," *BMC medical research methodology*, 10, 18.

Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Series B*, 34, 187220.

Davis, B. R. and Hardy, R. J. (1994), "Data monitoring in clinical trials: the case for stochastic curtailment," *Journal of clinical epidemiology*, 47, 1033–1042.

Demarqui, F. N., Loschi, R. H., and Colosimo, E. A. (2008), "Estimating the grid of time-points for the piecewise exponential model," *Lifetime data analysis*, 14, 333–356.

Dewé, W., Benoit, A., and Legrand, C. (2013), "Challenges and difficulties in running vaccine efficacy trials," *Expert Rev Pharmacoecon Outcomes Res*, 6, 403–411.

DiazGranados, C. A., Denis, M., and Plotkin, S. (2012), "Seasonal influenza vaccine efficacy and its determinants in children and non-elderly adults: a systematic review with meta-analyses of controlled trials," *Vaccine*, 31, 49–57.

Dimitrienko, A. and Wang, M.-D. (2006), "Bayesian predictive approach to interim monitoring in clinical trials," *Statistics in medicine*, 25, 2178–2195.

Duchateau, L. and Janssen, P. (2007), *The frailty model*, Springer.

Dunning, A. J. (2006), "A model for immunological correlates of protection," *Statistics in medicine*, 25, 1485–1497.

Edwards, K. M., Dupont, W. D., Westrich, M. K., Plummer, W. D., Palmer, P. S., and Wright, P. F. (1994), "A randomized controlled trial of cold-adapted and inactivated vaccines for the prevention of influenza A disease," *Journal of Infectious Diseases*, 169, 68–76.

Esposito, S., Marchisio, P., Cavagna, R., Gironi, S., Bosis, S., Lambertini, L., Droghetti, R., and Principi, N. (2003), "Effectiveness of influenza vaccination of children with recurrent respiratory tract infections in reducing respiratory-related morbidity within the households," *Vaccine*, 21, 3162–3168.

Ewell, M. (1996), "Comparing methods for calculating confidence intervals for vaccine efficacy," *Statistics in medicine*, 15, 2379–2392.

FDA (2007), "Guidance for industry: clinical data needed to support the licensure of seasonal inactivated influenza vaccines," *Rockville: US Department of Health and Human Services*.

Fine, P. E. (1993), "Herd immunity: history, theory, practice," *Epidemiologic Reviews*, 15, 265–302.

Flahault, A., Dias-Ferrao, V., Chaberty, P., Esteves, K., Valleron, A.-J., and Lavanchy, D. (1998), "FluNet as a tool for global monitoring of influenza on the Web," *Jama*, 280, 1330–1332.

Frome, E. (1983), "The analysis of rates using Poisson regression models," *Biometrics*, 665–674.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient parametrisations for normal linear mixed models," *Biometrika*, 82, 479–488.

Gelman, A. (2004), "Parameterization and Bayesian modeling," *Journal of the American Statistical Association*, 99.

Gilbert, P. B., Qin, L., and Self, S. G. (2008), "Evaluating a surrogate endpoint at three levels, with application to vaccine development," *Statistics in medicine*, 27, 4758–4778.

Grassly, N. C. and Fraser, C. (2006), "Seasonal infectious disease epidemiology," *Proceedings of the Royal Society B: Biological Sciences*, 273, 2541–2550.

Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T. M., Sutcliffe, J. G., and Lerner, R. A. (1982), "Immunogenic structure of the influenza virus hemagglutinin," *Cell*, 28, 477–487.

Gross, P. A., Hermogenes, A. W., Sacks, H. S., Lau, J., and Levandowski, R. A. (1995), "The efficacy of influenza vaccine in elderly personsA meta-analysis and review of the literature," *Annals of Internal medicine*, 123, 518–527.

Gupta, V., Earl, D. J., and Deem, M. W. (2006), "Quantifying influenza vaccine efficacy and antigenic distance," *Vaccine*, 24, 3881–3888.

Hahn, G. J. and Meeker, W. Q. (2011), *Statistical intervals: a guide for practitioners*, vol. 328, John Wiley & Sons.

Halloran, M. E., Haber, M., and Longini, I. M. (1992), "Interpretation and estimation of vaccine efficacy under heterogeneity," *American Journal of Epidemiology*, 136, 328–343.

Halloran, M. E., Haber, M., Longini, I. M., and Struchiner, C. J. (1991), "Direct and indirect effects in vaccine efficacy and effectiveness," *American Journal of Epidemiology*, 133, 323–331.

Halloran, M. E., Hayden, F. G., Yang, Y., Longini, I. M., and Monto, A. S. (2007), "Antiviral effects on influenza viral transmission and pathogenicity: observations from household-based trials," *American journal of epidemiology*, 165, 212–221.

Halloran, M. E., Longini, I. M., and Struchiner, C. J. (1996), "Estimability and interpretation of vaccine efficacy using frailty mixing models," *American Journal of Epidemiology*, 144, 83–97.

Halloran, M. E., Longini, I. M., Struchiner, C. J., Longini, I. M., and Struchiner, C. J. (2010), *Design and analysis of vaccine studies*, Springer.

Halloran, M. E., Struchiner, C. J., and Longini, I. M. (1997), "Study designs for evaluating different efficacy and effectiveness aspects of vaccines," *American journal of epidemiology*, 146, 789–803.

Haq, K. and McElhaney, J. E. (2014), "Immunosenescence: influenza vaccination and the elderly," *Current opinion in immunology*, 29, 38–42.

Harvey, B. S. and Kagerer, R. L. (2005), "bayesian predictive approach to interim monitoring in clinical trials," *Statistics in Medicine*.

Hay, A. J., Gregory, V., Douglas, A. R., and Lin, Y. P. (2001), "The evolution of human influenza viruses," *Philos Trans R Soc Lond B Biol Sci*, 356, 1861–70.

Heinze, G. and Schemper, M. (2001), "A solution to the problem of monotone likelihood in Cox regression," *Biometrics*, 57, 114–119.

— (2002), "A solution to the problem of separation in logistic regression," *Statistics in medicine*, 21, 2409–2419.

Higgins, J., Thompson, S. G., and Spiegelhalter, D. J. (2009), "A re-evaluation of random-effects meta-analysis," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 137–159.

Hoberman, A., Greenberg, D. P., Paradise, J. L., Rockette, H. E., Lave, J. R., Kearney, D. H., Colborn, D. K., Kurs-Lasky, M., Haralam, M. A., Byers, C. J., et al. (2003), "Effectiveness of inactivated influenza vaccine in preventing acute otitis media in young children: a randomized controlled trial," *Jama*, 290, 1608–1616.

Hobson, D., Baker, F., and Curry, R. (1973), "Effect of influenza vaccines in stimulating antibody in volunteers with prior immunity," *The Lancet*, 302, 155–156.

Hobson, D., Curry, R., Beare, A., and Ward-Gardner, A. (1972), "The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses," *Journal of Hygiene*, 70, 767–777.

Hoskins, T., Davies, J., Smith, A., Miller, C., and Allchin, A. (1979), "Assessment of inactivated influenza-A vaccine after three outbreaks of influenza A at Christ's Hospital," *The Lancet*, 313, 33–35.

Hougaard, P. (1999), "Fundamentals of survival data," *Biometrics*, 55, 13–22.

Hugo, J. (2012), "Bayesian Tolerance Intervals for Variance Component Models," Ph.D. thesis, University of the Free State.

ICH8 (1998), "General Considerations for Clinical Trials (ICH Topic E 8)," Tech. Rep. CPMP/ICH/291/95, Committee for Proprietary Medicinal Products.

Jackson, L. A., Gaglani, M. J., Keyserling, H. L., Balser, J., Bouveret, N., Fries, L., and Treanor, J. J. (2010), "Safety, efficacy, and immunogenicity of an inactivated influenza vaccine in healthy adults: a randomized, placebo-controlled trial over two influenza seasons," *BMC infectious diseases*, 10, 71.

Jennison, C. and Turnbull, B. W. (2010), *Group sequential methods with applications to clinical trials*, CRC Press.

John, T. J. and Samuel, R. (2000), "Herd immunity and herd effect: new insights and definitions," *European journal of epidemiology*, 16, 601–606.

Kawai, N., Ikematsu, H., Iwaki, N., Satoh, I., Kawashima, T., Tsuchimoto, T., and Kashiwagi, S. (2003), "A prospective, Internet-based study of the effectiveness and safety of influenza vaccination in the 2001–2002 influenza season," *Vaccine*, 21, 4507–4513.

Keitel, W. A., Cate, T. R., Couch, R. B., Huggins, L. L., and Hess, K. R. (1997), "Efficacy of repeated annual immunization with inactivated influenza virus vaccines over a five year period," *Vaccine*, 15, 1114–1122.

Kelly, M. J., Dunstan, F. D., Lloyd, K., and Fone, D. L. (2008), "Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods," *BMC psychiatry*, 8, 10.

Kim, J. S. and Proschan, F. (1991), "Piecewise exponential estimator of the survivor function," *Reliability, IEEE Transactions on*, 40, 134–139.

Koopman, P. (1984), "Confidence intervals for the ratio of two binomial proportions," *Biometrics*, 513–517.

Krishnamoorthy, K. and Mathew, T. (2009), *Statistical Tolerance Regions: Theory, Applications, and Computation*, vol. 744, John Wiley & Sons.

Lachin, J. M. (2011), *Biostatistical methods: the assessment of relative risks*, vol. 807, John Wiley & Sons.

Landis, J. R., Sharp, T. J., Kuritz, S. J., and Koch, G. G. (1998), "Mantel–haenszel methods," *Encyclopedia of Biostatistics*.

Lebrun, P. (2012), "Bayesian Design Space applied to Pharmaceutical Development," Ph.D. thesis, Université de Liège.

Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V., and Sylvester, R. (2009), "Validation of prognostic indices using the frailty model," *Lifetime data analysis*, 15, 59–78.

Li, C.-S., Taylor, J. M., and Sy, J. P. (2001), "Identifiability of cure models," *Statistics & Probability Letters*, 54, 389–395.

Li, S., Parnes, M., and Chan, I. S. (2013), "Determining the Cutoff Based on a Continuous Variable to Define Two Populations with Application to Vaccines," *Journal of biopharmaceutical statistics*, 23, 662–680.

Lin, D., Yao, Q., and Ying, Z. (1999), "A general theory on stochastic curtailment for censored survival data," *Journal of the American Statistical Association*, 94, 510–521.

Localio, A. R., Berlin, J. A., Ten Have, T. R., and Kimmel, S. E. (2001), "Adjustments for center in multicenter studies: an overview," *Annals of internal medicine*, 135, 112–123.

Lofgren, E., Fefferman, N., Naumov, Y., Gorski, J., and Naumova, E. (2007), "Influenza seasonality: underlying causes and modeling theories," *Journal of Virology*, 81, 5429–5436.

Longini, I. M., Koopman, J. S., Monto, A. S., and Fox, J. P. (1982), "Estimating household and community transmission parameters for influenza," *American Journal of Epidemiology*, 115, 736–751.

Longini Jr, I. M. and Halloran, M. E. (1996), "A frailty mixture model for estimating vaccine efficacy," *Applied Statistics*, 165–173.

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012), *The BUGS book: a practical introduction to Bayesian analysis*, CRC Press.

Machin, D., Cheung, Y. B., and Parmar, M. (2006), *Survival analysis: a practical approach*, John Wiley & Sons.

McElhaney, J. E., Beran, J., Devaster, J.-M., Esen, M., Launay, O., Leroux-Roels, G., Ruiz-Palacios, G. M., van Essen, G. A., Caplanusi, A., Claeys, C., et al. (2013), "AS03-adjuvanted versus non-adjuvanted inactivated trivalent influenza vaccine against seasonal influenza in elderly people: a phase 3 randomised trial," *The Lancet infectious diseases*, 13, 485–496.

McElhaney, J. E., Xie, D., Hager, W. D., Barry, M. B., Wang, Y., Kleppinger, A., Ewen, C., Kane, K. P., and Bleackley, R. C. (2006), "T cell responses are better correlates of vaccine protection in the elderly," *The Journal of Immunology*, 176, 6333–6339.

McNutt, L.-A., Wu, C., Xue, X., and Hafner, J. P. (2003), "Estimating the relative risk in cohort studies and clinical trials of common outcomes," *American journal of epidemiology*, 157, 940–943.

Michiels, S., Baujat, B., Mahe, C., Sargent, D., and Pignon, J. (2005), "Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses," *Journal of clinical epidemiology*, 58, 238–245.

Moghadas, S. (2004), "Modelling the effect of imperfect vaccines on disease epidemiology," *Discrete And Continuous Dynamical Systems Series B*, 4, 999–1012.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., et al. (2008), "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS medicine*, 5, e74.

Nardi, A. and Schemper, M. (2003), "Comparing Cox and parametric models in clinical studies," *Statistics in Medicine*, 22, 3597–3610.

Nauta, J. (2010), *Statistics in clinical vaccine trials*, Springer.

Nichol, K. L. (2006), "Heterogeneity of influenza case definitions and implications for interpreting and comparing study results," *Vaccine*, 24, 6726–6728.

O'Hagan, A., Stevens, J. W., and Campbell, M. J. (2005), "Assurance in clinical trial design," *Pharmaceutical Statistics*, 4, 187–201.

Ohmit, S. E., Victor, J. C., Rotthoff, J. R., Teich, E. R., Truscon, R. K., Baum, L. L., Rangarajan, B., Newton, D. W., Boulton, M. L., and Monto, A. S. (2006), "Prevention of antigenically drifted influenza by inactivated and live attenuated vaccines," *New England Journal of Medicine*, 355, 2513–2522.

Ohmit, S. E., Victor, J. C., Teich, E. R., Truscon, R. K., Rotthoff, J. R., Newton, D. W., Campbell, S. A., Boulton, M. L., and Monto, A. S. (2008), "Prevention of symptomatic seasonal influenza in 2005–2006 by inactivated and live attenuated vaccines," *Journal of Infectious Diseases*, 198, 312–317.

O'Quigley, J. and Stare, J. (2002), "Proportional hazards models with frailties and random effects," *Statistics in medicine*, 21, 3219–3233.

Park, T., Ki, M., and Yi, S.-G. (2004), "Statistical analysis of MMR vaccine adverse events on aseptic meningitis using the case cross-over design," *Statistics in medicine*, 23, 1871–1883.

Pawelec, G. (1999), "Immunosenescence: impact in the young as well as the old," *Mech Ageing Dev*, 108, 1–7.

Peng, Y. and Zhang, J. (2008), "Estimation method of the semiparametric mixture cure gamma frailty model," *Statistics in medicine*, 27, 5177–5194.

Pickles, A. and Crouchley, R. (1995), "A comparison of frailty models for multivariate survival data," *Statistics in Medicine*, 14, 1447–1461.

Qin, L., Gilbert, P. B., Corey, L., McElrath, M. J., and Self, S. G. (2007), "A framework for assessing immunological correlates of protection in vaccine trials," *Journal of Infectious Diseases*, 196, 1304–1312.

Reed, C., Meltzer, M. I., Finelli, L., and Fiore, A. (2012), "Public health impact of including two lineages of influenza B in a quadrivalent seasonal influenza vaccine," *Vaccine*, 30, 1993–1998.

Reed, W. J. (2011), "A flexible parametric survival model which allows a bathtub-shaped hazard rate function," *Journal of Applied Statistics*, 38, 1665–1680.

Rondeau, V., Filleul, L., and Joly, P. (2006), "Nested frailty models using maximum penalized likelihood estimation," *Statistics in medicine*, 25, 4036–4052.

Rothberg, M. B., Haessler, S. D., and Brown, R. B. (2008), "Complications of viral influenza," *The American journal of medicine*, 121, 258–264.

Sambhara, S. and McElhaney, J. E. (2009), "Immunosenescence and influenza vaccine efficacy," in *Vaccines for Pandemic Influenza*, Springer, pp. 413–429.

Senn, S. (2004), "Controversies concerning randomization and additivity in clinical trials," *Statistics in medicine*, 23, 3729–3753.

Skowronski, D. M., Tweed, S. A., and De Serres, G. (2008), "Rapid Decline of Influenza VaccineInduced Antibody in the Elderly: Is It Real, or Is It Relevant?" *Journal of Infectious Diseases*, 197, 490–502.

Smith, C. T., Williamson, P. R., and Marson, A. G. (2005), "Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes," *Statistics in medicine*, 24, 1307–1319.

Smith, P., Rodrigues, L., and Fine, P. (1984), "Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies," *International journal of epidemiology*, 13, 87–93.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995), "Bayesian approaches to random-effects meta-analysis: a comparative study," *Statistics in medicine*, 14, 2685–2699.

Struchiner, C. and Halloran, M. (2007), "Randomization and baseline transmission in vaccine field trials," *Epidemiology and infection*, 135, 181–194.

Taylor, J. M. (1995), "Semi-parametric estimation in failure time mixture models," *Biometrics*, 899–907.

Thomas, C. and Moridani, M. (2010), "Interindividual variations in the efficacy and toxicity of vaccines," *Toxicology*, 278, 204–210.

Tournoud, M. and Ecochard, R. (2008), "Promotion Time Models with Time-Changing Exposure and Heterogeneity: Application to Infectious Diseases," *Biometrical Journal*, 50, 395–407.

Treanor, J. J., Sahly, H. E., King, J., Graham, I., Izikson, R., Kohberger, R., Patriarca, P., and Cox, M. (2011), "Protective efficacy of a trivalent recombinant hemagglutinin protein vaccine against influenza in healthy adults: A randomized, placebo-controlled trial," *Vaccine*, 29, 7733–7739.

Truscott, J., Fraser, C., Cauchemez, S., Meeyai, A., Hinsley, W., Donnelly, C. A., Ghani, A., and Ferguson, N. (2011), "Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza," *Journal of The Royal Society Interface*, rsif20110309.

Tsai, R., Hotta, L. K., et al. (2013), "Polyhazard models with dependent causes," *Brazilian Journal of Probability and Statistics*, 27, 357–376.

Tsang, P., Gorse, G. J., Strout, C. B., Sperling, M., Greenberg, D. P., Ozol-Godfrey, A., DiazGranados, C., and Landolfi, V. (2014), "Immunogenicity and safety of Fluzone intradermal and high-dose influenza vaccines in older adults 65 years of age: A randomized, controlled, phase II trial," *Vaccine*, 32, 2507–2517.

Ulrichs, T. (2010), "Airborne Transmission: Influenza and Tuberculosis," in *Modern Infectious Disease Epidemiology*, Springer, pp. 279–290.

Weinstein, R. A., Bridges, C. B., Kuehnert, M. J., and Hall, C. B. (2003), "Transmission of influenza: implications for control in health care settings," *Clinical Infectious Diseases*, 37, 1094–1101.

WHO (2004), "Guidelines on clinical evaluation of vaccines: regulatory expectations," WHO Technical Report 924, WHO.

— (2012), "Vaccines against influenza WHO position paper-November 2012," *Weekly Epidemiol Rec*, 87, 461–76.

Wienke, A. (2010), *Frailty models in survival analysis*, CRC Press.

Wolfinger, R. D. (1998), "Tolerance intervals for variance component models using Bayesian simulation," *Journal of Quality Technology*, 30, 18–32.

Youden, W. J. (1950), "Index for rating diagnostic tests," *Cancer*, 3, 32–35.

Zhang, J. and Kai, F. Y. (1998), "What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes," *Jama*, 280, 1690–1691.

Zou, K. H., OMalley, A. J., and Mauri, L. (2007), "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, 115, 654–657.

Zweig, M. H. and Campbell, G. (1993), "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine." *Clinical chemistry*, 39, 561–577.